AN ANALYSIS OF VARIABLES AFFECTING INSTRUCTIONAL EFFICIENCY

A thesis submitted in partial fulfilment

of the requirements for the Degree

of

Doctor of Philosophy

in the

University of Canterbury

by

Kyle G. M^c^Williams

University of Canterbury

2006

CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURES

ACKNOWLEDGMENTS

This thesis was made possible because of the support of many people. Foremost was the support from my supervisor, Dr. John Church, who expertly shaped my behaviour. John's patient guidance and encouragement, and his sense of humour are greatly appreciated. Dr. Kathleen Liberty was a joint supervisor for much of the thesis, and for her advice I am very thankful.

I am indebted to Mr. Bruce Hodgson and Dr. Neil Averis. Without the help, support and belief of these two gentlemen I would not have been able to begin this thesis.

I would like to thank the principal for allowing access to the school, to the teachers and the students, and for providing me with any required resources. I greatly appreciate the willingness of the teachers who made their students available to participate, despite any disruptions this may have caused. Finally, I wish to thank all the children who volunteered for the experiments for their enthusiasm in participating.

Kyle G. McWilliams

(September, 2005)

ABSTRACT

A lot about the learning process still remains unknown. The experiments described in this thesis investigated variables that affect instructional efficiency by employing specifically programmed computers to manage and control instructional variables within each experiment for 6- to 7-year old children.

A Measurement Procedures Study was undertaken to ascertain when a response should be classified as "acquired." It was decided to classify a response as acquired if it could be performed correctly (without prompting) seven days after instruction.

A review of the relationship between accuracy level during instruction and the rate of acquisition found that higher accuracy levels during instruction tend to be associated with higher rates of acquisition provided that non-copying prompting procedures are employed.

The first experiment investigated the relationship between accuracy level during instruction and rate of acquisition by presenting a non-copying antecedent prompt (model of the correct spelling word) depending on a preselected target accuracy level. As an error-contingent prompt (model of the correct spelling word) was also provided it could not be ascertained whether transfer of stimulus control occurred as a result of the antecedent prompt, or the error-contingent prompt, or both. The second experiment was a repeat of the first experiment with the error-contingent prompt removed. It was found that it was possible to manage, although not completely control, the accuracy level during instruction by presenting a simultaneous non-copying prompt and that higher accuracy levels during instruction were associated with higher rates of acquisition.

A review that examined the error-correction research found that a variety of correction procedures were effective. However, none of the 36 experiments which were reviewed controlled the number of response opportunities.

Experiment 3 compared the effects on rate of acquisition of presenting an antecedent

model or an error-contingent model. The results of Experiment 3 showed that when the number of learning opportunities was controlled there was little difference in effectiveness or efficiency between an antecedent model and an error-contingent model.

Experiment 4 compared the effects of presenting an error-contingent model against an error-contingent model and a secondary response opportunity. It was found that an error-contingent model was at least as effective, although it was overall less efficient when response opportunities were controlled.

A supplementary analysis was undertaken to review and compare the results obtained across the four experiments. Across experiments each newly acquired spelling response required about five practice responses, on average. It appears this was a critical variable for acquisition. Additionally, each acquired response was acquired over a two-day period. Although rates of acquisition differed between high-achieving children and low-achieving children, there was little difference in the number of practice responses required for acquisition between these two groups. It was observed that most of the 6- to 7-year old participants found error feedback aversive and this appeared to result in reduced attention to models of the correct spelling when these occurred following errors. The results from this series of investigations suggest that an opportunity for the transfer of stimulus control from the prompt (model of the correct spelling) to the practice stimulus (the spoken word) is more critical for acquisition than where the prompt occurs within the trial (that is, the antecedent or consequent position).

It was suggested that future research could investigate (a) the variables which are necessary for the transfer of stimulus control, (b) the generality of the observation that children require five practice responses in order to acquire discrete academic responses, and (c) the effects on rates of acquisition and instructional efficiency of varying the distribution in time of practice responses for children who are learning various types of academic skills.

CHAPTER 1

INTRODUCTION

If Ebbinghaus (1885/1913) is considered to be the first to investigate human learning and remembering then the human learning process has been the subject of intensive research for more than a century. In spite of this research effort, there is still much about the learning process that we do not understand. It is not surprising, therefore, that there is much about the process of facilitating learning (the process of teaching) that we also do not understand. “Few facts concerning teacher effectiveness have been established” (Ornstein, 1991, p. 63). The basic shape of a classroom is the same as it was 100 years ago. “Little has changed in the ways that schools divide time and space, classify students and allocate them to classrooms, splinter knowledge in to “subjects,” and award grades and “credits” as evidence of learning” (Tyack & Cuban, 1995, p. 85). The great majority of teacher education programmes continue to teach student teachers a range of theories of learning (Gage & Berliner, 1998; Hill, 1977), to communicate a range of views about how learning is best encouraged in children, and to encourage student teachers to develop their own unique teaching philosophy and teaching style. This gives student teachers the impression that specific instructional variables such as feedback are less important than personality characteristics such as encouragement and enthusiasm.

What is Learning?

Learning has been defined in a variety of ways. Gage and Berliner (1998, p. 208) define learning as the “process whereby an organism changes its behavior as a result of experience.” Cognitive scientists such as Stillings et al. (1987, p. 189) define learning as “any process whereby people increase their knowledge or improve their skill.” Behaviour analysts such as Catania (1998, p. 395) define learning as “the process by which behavior is added to an organism's repertory; a relatively permanent change in behavior." In each of these definitions, learning is defined as a process. "The idea that learning is a *process* means that learning takes time. To measure learning, we compare the way an organism behaves at time 1 with the way it behaves at time 2 under similar circumstances. If the behavior under similar circumstances differs on the two occasions, we infer that learning has taken place" (Gage & Berliner, 1998, p. 208).

Theoretical Approaches to the Analysis of Learning Processes

For the past 50 years the experimental analysis of learning processes has been undertaken within two separate and largely incompatible theoretical orientations; a cognitive orientation and a behaviour analytic orientation. Cognitive theories of learning are attempts to build structural rather than functional models of the learning process. That is, investigations are carried out in an attempt to identify relationships between overt performance and mental processing. Performance is assumed to be a function of previously acquired mental structures. Cognitive theorists build models of mental structures which may operate to regulate learning and remembering in learners (Stillings et al., 1987). The purpose of developing cognitive theories of learning is to "establish functional relationships among phenomena with a view to predicting and, if possible, to controlling their occurrence" (Mouly, 1970, p. 39). There are generally two types of explanations of learning in cognitive science; dispositional explanations, and mechanistic explanations. Dispositional explanations invoke internal traits to explain differences in the way individuals respond to external events with the internal traits bridging the gap between experience and performance. For example, differences in performance may be attributed to differences in intelligence, ability, self-esteem, self-efficacy and so on. Mechanistic explanations on the other hand use some form of mental structure or cognitive processing variable to bridge the gap between experience and performance. "The overall goal of instruction is to help students construct mental representations that correctly or accurately mirror … relationships located outside the mind in instructional representations" (Cobb, Yackel, & Wood, 1992, p. 4). For example, some explanatory models are models of memory where internal representations of knowledge are stored in memory. Knowledge is transferred from the external world via the senses to short-term memory (working memory) where it is worked on or adapted as a result of prior knowledge and is then transferred to a long-term memory store where it can be retrieved on future occasions by the short-term or working memory (Nuthall, 2000a).

The theoretical orientation of behaviour analysts is functional rather than structural. Behaviour analysts ask what a behaviour accomplishes rather than what it looks like and how is it represented in memory (Baum, 1994). "The interest in learning about how certain environmental variables may influence the response class lies in the belief that behavior is largely the result of environmental variables" (Johnston & Pennypacker, 1993a, p. 11). Behaviour analysts seek to develop theories of behaviour change that permit both prediction and control by identifying functional relationships between a particular type of behaviour change and the environmental variables that might be necessary for that behaviour change to occur. Behaviour analysts group responses together based on their common purpose or function because behaviours that look the same may serve different functions. For example, handing a wallet to a friend is different from handing a wallet to a mugger (Baum, 1994). The aim of behaviour analysis research is to improve our ability to "predict and control the behavior of the individual organism" (Skinner, 1953, p. 35). This is considered to be a realistic research agenda because thousands of controlled observations have shown that responses which result in outcomes favourable to the learner, tend to be used again in the future.

Phases of Learning

Although cognitive scientists and behaviour analysts approach learning from different perspectives both agree that there are several "phases" of learning. The cognitive scientist Anderson (1985) identifies three phases of learning. The first is the cognitive phase where the learner knows the skill that needs to be learned. The second is the associative phase where a method for performing the skill is worked out and the third, the autonomous phase, is where the skill becomes more and more rapid and automatic. Nuthall (2000b) argues that sensations enter via the sensory register and are transferred to the working memory where they are stored. They are then either (a) forgotten if not joined to other relevant experiences, or (b) transferred to long-term memory. "In a typical classroom context it takes 3 -4 experiences for the processes of connecting, elaborating, and integrating to produce a new concept that is transferred to long-term memory" (Nuthall, 2000b, p.7). With further practice, new skills and abilities become automatised. "Automatic processing can be defined as a rapid, accurate type of processing that requires minimal awareness and attention" (Podell, Tournaki-Rein, & Lin, 1992, p. 200). It is thought that automaticity of a response is a function of the amount of practice that a learner receives. However, cognitive scientists have shown little interest in research designed to identify the amount of practice required in order to "automatise" different kinds of knowledge and skills.

Behaviour analysts also distinguish between several phases of learning. Johnson and Layng (1994) describe four phases. The first phase they refer to as "establishing" the response. During the establishing phase the learner learns to respond correctly to instructional stimuli. The next phase is the development of fluency. The learner learns to respond not only accurately but also quickly or proficiently. The next phase is the development of endurance. During this phase the learner learns to respond fluently for longer periods of time. The fourth phase, the application phase, is where the learner learns to apply the new skills to new environments and to combine them with more complex skills. Rivera and Smith (1997) identify five phases; acquisition, proficiency, maintenance, generalisation, and adaption. The establishment of the acquisition phase is marked by a change from the inability of the learner to respond correctly to the ability to respond correctly without prompting.

The experimenter who seeks to study learning must specify the phase or phases of learning that are to be investigated because different variables assume critical importance during each of the different phases. For example, differential reinforcement for responding correctly is critical during the acquisition phase (e.g., Trap, Milner-Davis, Joseph, & Cooper, 1978) whereas differential reinforcement for increases in rate of correct responding is critical during the fluency building phase (e.g., Noell et al., 1998; Sulzer, Hunt, Ashby, Koniarski, & Krams, 1971).

The experiments described in this thesis examine the effects of variables that operate during the acquisition phase. This thesis focuses on initial acquisition because we still have not discovered the key determinants of efficient learning during the acquisition phase.

Variables Thought to Influence Rate of Acquisition

The rate at which children acquire new skills and understandings is thought to be affected by a number of variables. These variables include (a) the appropriateness of the teaching aim, (b) the opportunity to respond, (c) the level of fluency required before moving to the next teaching aim, (d) the procedure used to monitor individual student progress, and (e) the appropriateness of the instructional procedures employed.

Appropriateness of the Teaching Aim

An appropriate teaching aim is one that takes the learner’s current skill level into account. “A knowledge of what the learner can do already is a critical element in teaching because it is the one factor, more than any other, which determines what can be learned (and taught) next” (Church, 1999a). Only by accurately identifying what a learner can and cannot do prior to instruction can a teacher ensure that practice is (a) not so easy that the learner is simply practising responses which they have already acquired and (b) not so difficult that the learner loses motivation, interest and self-confidence (e.g., Schumm, Moody, & Vaughn, 2000).

Opportunity to Respond

Educational researchers have employed a variety of measures of student engagement. Academic learning time (C. S. Fischer et al., 1980) is the total time that the learner engages with the practice material and has been found to be strongly correlated with achievement (C. W. Fischer & Berliner, 1985; Rosenshine & Berliner, 1978). However, academic learning time does not take into account the number of learning trials a student engages in during instruction and it is this, not the passage of time, that affects learning (Greenwood, Delquadri, & Hall, 1984). Opportunity to respond refers to the occasions when the learner gets to practise the to-be-learned responses or skills (Hall, Delquadri, & Harris, 1977). Response opportunities can be counted. Counts of response opportunities can be used to measure how much instruction students are receiving (Heward, 1994). Counting practice responses also enables the researcher to measure the number of trials which are required to learn new responses in the classroom.

Level of Fluency or Automaticity Required Before Moving to the Next Teaching Aim

Skills that are practised only to acquisition are often forgotten. However, skills practised to high levels of fluency (automaticity) tend to be maintained for long periods of time (Freeman & Haughton, 1993; Hasbrouck & Tindal, 1992). The level of fluency (speed of recall) that a learner achieves is a good predictor of maintenance (long-term retention). In order to achieve long-term retention Binder, Haughton and Bateman (2002) suggest practising to a fluency level of 60-80 correct responses per minute for naming-type responses. While fluency is important in the learning process, it is not relevant in this thesis as the present series of experiments investigates variables that are important during the acquisition phase of learning.

Procedure Used to Monitor Individual Student Progress

Another variable which influences rate of acquisition is the procedure used to monitor individual student progress. If the teacher bases teaching decisions on data regarding the learner's rate of improvement, this may affect rate of acquisition (Stecker & Fuchs, 2000). Curriculum-Based Measurement systems (Deno, 1985; Deno & Fuchs, 1987) measure individual student achievement in curriculum areas by sampling curriculum objectives using 2- to 5-minute tests administered every day or two. This data is then used to adjust teaching methods and learning activities for individual children. The introduction of Curriculum-Based Measurement systems have been shown to accelerate progress in at least some children (Stecker & Fuchs, 2000).

Appropriateness of the Instructional Procedures Employed

Different instructional procedures may be used in the classroom. The instructional procedure that is appropriate depends upon the type of learning outcome that the teacher is aiming for (Church, 1999b). During the acquisition phase different kinds of learning outcomes are possible. Gagne (1977) distinguished between a number of different learning outcomes including discriminations and concrete concepts, defined concepts and rules, cognitive problem solving, verbal learning, motor skills, and attitudes. Engelmann and Carnine (1991) describe different teaching procedures for chained tasks, concepts, discriminations, and relationships. Kameenui and Simmons (1990) distinguish between the learning of new facts and discriminations, concepts, rule relations, reading decoding, reading comprehension, mathematical facts, concepts and operations, and expressive writing. Of particular interest in this thesis are the instructional procedures used by the teacher during the acquisition phase when children are learning discrete one-to-one matching responses such as new spelling responses. With younger children, two of the most important instructional variables during this type of learning appear to be prompting and feedback.

Prompting

Of the many variables that affect rate of acquisition, presenting the learner with a stimulus that already has some stimulus control over the correct response appears to be critical. The word *control* is used to refer to exerting an influence over behaviour. "It is important to recognize . . . that antecedent events influence operant behavior in conjunction with consequent events" (Vollmer & Van Camp, 1998, p. 96). Stimuli that have already acquired stimulus control are often referred to as discriminative stimuli (S^D). An S^D can be presented in either the antecedent position as a prompt or the consequent position in the form of a correction following an error.

Both cognitive scientists and behaviour analysts talk about providing assistance prior to the learner responding. Cognitive scientists often use the term scaffolding (Wood, Bruner, & Ross, 1976). Scaffolding is a "process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts" (Wood et al., 1976, p. 90). Modelling is a specific form of scaffolding where learning "takes place as a result of seeing someone else carry out the performance" (Biggs & Moore, 1993, p. 527).

Behaviour analysts use the term prompting. A prompt is an antecedent stimulus presented to a learner that increases the probability of a correct or improved response. This usually involves showing or telling the learner how to respond (Cooper, Heron, & Heward, 1987). Showing is usually referred to as modelling. Modelling is "demonstrating the desired behavior to one or more observers. The observers are then required to imitate the skill" (Rivera & Smith, 1997, p. 244). When presenting a prompt following an instructional stimulus the aim is to transfer stimulus control from the prompt to the instructional stimulus.

Different prompting procedures are possible. Delayed prompting involves a delay between the presentation of the instructional stimulus and the presentation of the prompt. This provides the learner with the opportunity to respond correctly prior to the presentation of the prompt (Handen & Zane, 1987). Most-to-least prompting involves presenting progressively less intrusive prompts as the learner demonstrates that they can perform the response (Wolery, Bailey, & Sugai, 1988). This compares to the system-of-least prompting (or least-to-most) which involves presenting progressively stronger and stronger prompts until the learner responds correctly (Gast, Ault, Wolery, & Doyle, 1988).

Feedback

Both cognitive scientists and behaviour analysts recognise the importance of consequences (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Grant & Evans, 1994). One of the most important consequences during acquisition is feedback. Cognitive scientists define feedback in learning as “any information about the correctness or appropriateness of a response” (Reber, 1995, p. 283). This definition is sometimes extended to include consequences “which serve as sources of new information necessary for verification of retrieval accuracy, concept development, skill refinement, and metacognitive adaptation” (Bangert-Drowns et al., 1991, p. 214). Because it occurs as a consequence, feedback can sometimes function as a reinforcer and sometimes function as a punisher. However, feedback is not defined in terms of its reinforcing or punishing properties (Grant & Evans, 1994). Catania (1998, p. 390) defines feedback as “a stimulus or stimulus properties correlated with or produced by the organism’s own behavior.” Feedback in this thesis will be defined as a stimulus presented to a learner, contingent upon a response, that indicates whether the response is correct or incorrect. The natural environment, another person, or a computer may provide this signal. Prompts and other forms of assistance can also be provided in the consequence position - especially following errors. When a learner makes an error, presenting a prompt in the consequent position is often referred to as error correction (Barbetta, Heron, & Heward, 1993; Grimes, 1981). However, it might more appropriately be referred to as an error-contingent prompt and this is the term that will be used in this thesis.

While there are a number of variables that influence rate of acquisition, it appears that two of these, prompting and feedback, are critical. Prompting and feedback variables were therefore selected for analysis in the present series of experiments.

Measuring Acquisition, Rate of Acquisition, and Instructional Efficiency

Acquisition

Acquisition has been measured in many different ways. Cognitive scientists often refer to postexperimental retention. However, retention intervals vary widely across experiments from end-of-session retention (e.g., Elley, 1966) through to 12-month retention (e.g., Nuthall, 2000a). Behaviour analysts also measure acquisition in a variety of ways with tests of acquisition occurring anywhere from end-of-session (e.g., Barbetta, Heward, & Bradley, 1993) through to several weeks following instruction (e.g., Okyere, Heron, & Goddard, 1997). For the purposes of the present experiments, acquisition was initially defined as a change from the learner being unable to respond correctly to being able to respond correctly without prompting at least seven days after instruction.

Rate of Acquisition

Rate of acquisition is the cumulative number of correct responses acquired per unit of practice. The unit of practice is likely to be one “session.” Session length can be measured in either time or trials. There is little agreement as to how long the session should be when it is measured in time. However, this is usually solved pragmatically by having a session length that is appropriate for the children of a particular age. For example, Grskovic and Belfiore (1996) reported the cumulative number of responses acquired per 30-minute session with five 10- and 11-year old participants classified as emotionally handicapped. A problem with measuring the unit of practice (session length) in time is that participants respond at different speeds with the result that individual participants experience different amounts of practice (different numbers of trials) per session.

An alternative to measuring the session length in time is to measure the number of trials. However, there is little agreement as to the number of trials that a session should contain. Again, this is solved pragmatically by having an age-appropriate number of trials in a session. For example, Barbetta, Heward and Bradley (1993) reported the cumulative number of responses acquired per 30-trial session for four 8- to 9-year olds classified as developmentally delayed, while Mechling and Langone (2000) reported the cumulative number of responses acquired per 15-trial session for an 11- and 24-year old classified as severely intellectually disabled. An advantage of measuring the session length in number of trials is that this is the variable which affects rate of learning, not the amount of time (Greenwood et al., 1984).

Rate of acquisition depends upon (a) the response size, (b) the session length, and (c) the instructional variables. To measure the effects of instructional variables on rate of acquisition both the response size and the session length must be controlled. Response size affects rate of acquisition because practising smaller responses (e.g., unknown 3-letter spelling words) for a given period of time produces higher rates of acquisition than acquiring larger responses (e.g., unknown 6-letter spelling words) for the same time period. Controlling response size and the session length can be achieved (a) by selecting tasks where the responses differ little in size and effort, and (b) by having the same number of trials in each practice session. For the experiments described in this thesis it was decided to measure rate of acquisition as the cumulative number of responses acquired per given number of trials. Thus defined, differences in rate of acquisition appear as differences between treatments in the slope of the cumulative responses acquired from session to session.

Instructional Efficiency

A sequence of instruction may be more or less efficient. Given that there are a limited number of instructional hours in the school day, it is important that teachers use this time wisely. In this thesis instructional efficiency is defined in terms of learner effort (Saunders & Saunders, 1998). An instructional procedure is said to be more efficient if it results in the child learning to respond correctly to instructional stimuli with less effort than required during another instructional procedure. Effort may be measured in terms of the mental or physical effort required to produce a correct response or in terms of the number of practice responses required (Friman & Poling, 1995; Wolery, Ault, & Doyle, 1992). For example, a teaching procedure where a child requires 5 practice trials before learning and remembering how to spell the word *mother* is a more efficient procedure than one that results in a child requiring 10 practice trials before learning and remembering how to spell the word *mother*.

The pretest-posttest procedure cannot be used to measure instructional efficiency because the point at which a response is acquired cannot be detected. Instructional efficiency can, however, be measured by counting trials to criterion. Trials to criterion is measured by counting the number of practice trials on a response required in order to acquire that response. For example, a learner might require 10 practice trials on the word *school* in order to acquire the spelling of the word *school*. The instructional efficiency would therefore be 10 trials to criterion. This instructional efficiency procedure has two main advantages. First, it is sensitive enough to measure (a) physical effort at the trial level and (b) the point at which a response is acquired. Second, as there is no agreed unit of practice, it allows comparisons between treatments and experiments to be made where the session lengths differ. Given these advantages, trials to criterion has been selected as the measure of instructional efficiency in this series of experiments.

Measuring the Effects of Teaching on Learning

Analyses of learning are currently undertaken using two different experimental procedures; between-groups procedures and within-subject procedures. In between-groups experiments, different participants experience different treatments - usually an experimental treatment and a control treatment. The mean posttreatment performance of participants in the experimental treatment is compared to the mean posttreatment performance of participants in the control treatment. In within-subject experiments, each participant experiences both the experimental and the control treatment, and the performance of each individual participant during the experimental treatment is compared against their performance during the control treatment (Poling, Methot, & LeSage, 1995).

Behaviour is an individual phenomenon (Johnston & Pennypacker, 1993b) and learning researchers should attempt to explain and predict the learning of individuals. A major weakness of the between-groups procedure is that mean performance data provide no information on how individual participants responded. That is, a mean posttreatment score is rarely representative of every participant in that treatment and knowing "that the average performance of a group changed in a given way tells little about the performance of individual subjects. It is quite possible that the average performance of subjects in the experimental group improved while the performance of some subjects stayed the same and the performance of others deteriorated" (Cooper et al., 1987, p. 232). Within-subject procedures measure the performance of individuals. This allows the variables responsible for a change in individual performance to be identified.

In order for an experiment to generate data of scientific value it must provide an accurate and reliable measure of the effects of the independent variable on learning. The only way of ascertaining whether or not an independent variable has a reliable effect on learning is to repeat the experiment under similar conditions. That is, "Will the experiment, if repeated, yield the same results?" (Sidman, 1960, p. 43). In the present experiments replication was achieved by repeating the experiment across individual subjects.

One source of variability that participants always bring to an experiment is their prior learning history. Participants respond in ways that are partially the result of prior experience, and this responding may or may not be relevant in the experiment (Johnston & Pennypacker, 1993b). Variability due to prior learning history in between-groups experiments is attributed to chance. However, if "such variables are in fact controllable, then chance in this sense is simply an excuse for sloppy experimentation" (Sidman, 1960, p. 45). Attempts to cancel out variability by statistical manipulation "neither eliminates its presence in the data nor controls the function of the variables responsible for it" (Cooper et al., 1987, p. 233). Barlow and Hersen (1984) argue that this lack of control in between-groups designs is often the source of weak results. In the present experiments this problem was avoided by using within-subject designs in which each learner served as his or her own control.

In order to obtain an accurate measure of the effects of specific teaching variables on rate of acquisition it is important that the experiment is designed in such a way as to rule out alternative explanations for any difference in the rate of acquisition observed following the introduction of the independent variable (Johnston & Pennypacker, 1993b; Poling et al., 1995). This is achieved by exercising a high degree of experimental control over the non-experimental variables which are likely to affect acquisition. Sometimes this can be achieved by eliminating extraneous variables although this is oftentimes practically impossible. An alternative is for the experimenter to hold extraneous variables constant across treatments. While this does not eliminate their effects on the dependent variable it helps reduce the likelihood of their effects being confounded with those of the independent variable (Poling et al., 1995). The main extraneous variables controlled during the present experiments were the number of trials per session, the size of the practice set, the number of presentations of each practice stimulus per session, and standardising the presentation of feedback.

Learning Task

It is important to select a learning task which can be held constant from one treatment to the next and from one experiment to the next (Barlow & Hersen, 1984). This means that the learning researcher must select a learning task which involves only a single type of learning outcome so that individual responses share "common determinants in the surrounding environment" (Johnston & Pennypacker, 1993b, p. 188).

It was decided to select spelling as the learning task in this series of experiments because (a) there is a single type of learning outcome, (b) the degree of variability in spelling responses is small and the response size can be controlled, and (c) spelling is a socially valid task that is part of the classroom programme for 6- to 7-year old children. In addition, it is possible to manage the task difficulty of a spelling response by counting letter-pairs of a word (White & Haring, 1980). For example, the word *dog* has four letter-pairs; the space before the *d* and the *d*, the *d* and the *o*, the *o* and the *g*, and the *g* and the space following. Using this calibration it is possible to count (a) the number of letter-pair errors and (b) the total letter-pairs of a word prior to an experiment. This allows the experimenter to assign words to treatments based on the degree of difficulty rather than random assignment.

Summary

The present experiments sought to identify instructional variables that affect rate of acquisition and instructional efficiency during the acquisition phase of learning on spelling tasks with young children. Rate of acquisition was to be measured by counting the cumulative number of responses acquired per unit of practice. Instructional efficiency was to be measured by counting trials to criterion. The effects of instructional variables on rate of acquisition were measured by employing within-subject designs.

REFERENCES

Anderson, J. C. (1985). *Cognitive psychology and its implications (2nd ed.)*. New York: W. H. Freeman and Co.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., Heward, W. L., & Bradley, D. M. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110.

Barlow, D. W., & Hersen, M. (1984). *Single case experimental designs*. New York: Pergamon Press.

Baum, W. M. (1994). *Understanding behaviorism: Science, behavior, and culture*. New York: HarperCollins College Publishers.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning (3rd ed.)*. New York: Prentice Hall.

Binder, C., Haughton, E., & Bateman, B. (2002). *Fluency: Achieving true mastery in the learning process*. Retrieved January 29, 2004, from http://www.haughtonlearningcenter.com

Catania, A. C. (1998). *Learning (4th ed.)*. New Jersey: Prentice Hall.

Church, R. J. (1999a). *Basic learning processes*: University of Canterbury, Department of Education.

Church, R. J. (1999b). *Instructional processes*: University of Canterbury, Department of Education.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 23*, 2-33.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus: Merrill Publishing Company.

Deno, S. L. (1985). Curriculum-based measurement. *Exceptional Children, 52*, 63-71.

Deno, S. L., & Fuchs, D. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University (Original work published 1885, translated 1913).

Elley, W. (1966). The role of errors in learning with feedback. *Journal of Educational Psychology, 36*, 296-300.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and application*. Eugene: ADI Press.

Fischer, C. S., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement. In C. Denham & A. Lieberman (Eds.), *Time to learn*. Washington, DC: National Institute of Education.

Fischer, C. W., & Berliner, D. C. (1985). *Perspectives on instructional time*. New York: Longman.

Freeman, G., & Haughton, E. (1993). Building reading fluency across the curriculum. *Journal of Precision Teaching, 10*, 29-30.

Friman, P. C., & Poling, A. (1995). Making life easier with effort: Basic findings and applied research on response effort. *Journal of Applied Behavior Analysis, 28*, 583-590.

Gage, N. L., & Berliner, D. C. (1998). *Educational psychology (6th ed.)*. Boston: Houghton Mifflin.

Gagne, R. M. (1977). *The conditions of learning* (3rd ed.): Holt, Rinehart and Winston.

Gast, D. L., Ault, M. J., Wolery, M., & Doyle, P. M. (1988). Comparison of constant time delay and the system of least prompts in teaching sight word reading to students with moderate retardation. *Education and Training in Mental Retardation, 23*, 117-128.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: HarperCollins.

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1984). Opportunity to respond and student academic performance. In W. L. Heward, T. E. Heron, D. S. Hill & J. Trap-Porter (Eds.), *Focus on Behaviour Analysis in Education*. Ohio: Charles E. Merrill Publishing Company.

Grimes, L. (1981). Error analysis and error correction procedures. *Teaching Exceptional Children, 14*, 17-21.

Grskovic, J. A., & Belfiore, P. J. (1996). Improving the spelling performance of students with disabilities. *Journal of Behavioral Education, 6*, 343-354.

Hall, R. V., Delquadri, J. C., & Harris, J. (1977). *The opportunity to respond: A new focus for applied behavior analysis.* Paper presented at the Annual Meeting of the Association for Behavior Analysis, Chicago.

Handen, B. L., & Zane, T. (1987). Delayed prompting: A review of procedural variations and results. *Research in Developmental Disabilities, 8*, 307-330.

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in Grades 2 through 5. *Teaching Exceptional Children, 24*, 41-44.

Heward, W. L. (1994). Three "low tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283-320). California: Brooks/Cole Publishing Company.

Hill, W. F. (1977). *Learning: A survey of psychological interpretations (3rd ed.)*. Bungay: Richard Clay (The Chaucer Press) Ltd.

Johnson, K. R., & Layng, T. V. (1994). The morningside model of generative instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction*. California: Brooks/Cole Publishing Company.

Johnston, J. M., & Pennypacker, H. S. (1993a). *Readings for strategies and tactics of behavioral research* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Johnston, J. M., & Pennypacker, H. S. (1993b). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Kameenui, E. J., & Simmons, D. C. (1990). *Designing instructional strategies: The prevention of academic learning problems*. Ohio: Merrill Publishing Company.

Mechling, L., & Langone, J. (2000). The effects of a computer-based instructional program with video anchors on the use of photographs for prompting augmentative communication. *Education and Training in Mental Retardation and Developmental Disabilities, 35*, 90-105.

Mouly, G. J. (1970). *The science of educational research* (2nd ed.). New York: Van Nostrand Reinhold Company.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., et al. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31*, 659-663.

Nuthall, G. (2000a). The role of memory in the acquisition and retention of knowledge in science and social studies units. *Cognition and Instruction, 18*, 83-139.

Nuthall, G. (2000b). *Understanding how classroom experience shapes student's minds.* Paper presented at the Kongress der Deutschen Gessellschaft fur Erziehungswissenschaft, Gottingen.

Okyere, B. A., Heron, T. E., & Goddard, Y. (1997). Effects of self-correction on the acquisition, maintenance, and generalization of the written spelling of elementary school children. *Journal of Behavioral Education, 7*, 51-69.

Ornstein, A. C. (1991). Teacher effectiveness research: Theoretical considerations. In H. C. Waxman & H. J. Walberg (Eds.), *Effective teaching: Current research*. Baltimore: Paul H. Brookes Publishing Co.

Podell, D. M., Tournaki-Rein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted instruction among students with mild mental handicaps. *Education and Training in Mental Retardation, 27*, 200-206.

Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum Press.

Reber, A. S. (1995). *Dictionary of Psychology (2nd ed.)*. London: Penguin Books.

Rivera, D. P., & Smith, D. D. (1997). *Teaching students with learning and behavior problems (3rd ed.)*. Boston: Allyn and Bacon.

Rosenshine, B., & Berliner, D. C. (1978). Academic engagement time. *British Journal of Teacher Education, 4*, 3-16.

Saunders, R. R., & Saunders, M. D. (1998). Supported routines. In J. K. Luiselli & M. J. Cameron (Eds.), *Antecedent control:Innovative approaches to behavioral support* (pp. 245-272). Baltimore: Paul H. Brookes Publishing Co.

Schumm, J. S., Moody, S., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities, 33*, 477-488.

Sidman, M. (1960). *Tactics of scientific research*. Waltham: Blaisdell.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement. *Learning Disabilities Research & Practice, 15*, 128-134.

Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenbaum, D. A., Weisler, S. E., et al. (1987). *Cognitive science: An introduction*. Massachusetts: Massachusetts Institute of Technology.

Sulzer, B., Hunt, S., Ashby, E., Koniarski, C., & Krams, M. (1971). Increasing rate and percentage correct in reading and spelling in a fifth grade public school class of slow readers by means of a token system. In E. A. Ramp & B. I. Hopkins (Eds.), *A new direction for education: Behavior analysis* (Vol. 1, pp. 5-28). Kansas: University of Kansas.

Trap, J. J., Milner-Davis, P., Joseph, S., & Cooper, J. O. (1978). The effects of feedback and consequences on transitional cursive letter formation. *Journal of Applied Behavior Analysis, 11*, 381-393.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Massachusetts: Harvard University Press.

Vollmer, T. R., & Van Camp, C. M. (1998). Experimental designs to evaluate antecedent control. In J. K. Luiselli & M. J. Cameron (Eds.), *Antecedent control: Innovative approaches to behavioral support*. Baltimore: Paul H. Brookes Publishing Co.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching (2nd ed.)*. Ohio: Charles E. Merrill Publishing Company.

Wolery, M., Ault, M. J., & Doyle, P. M. (1992). *Teaching students with moderate to severe disabilities*. New York: Longman.

Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional students*. Massachusetts: Allyn and Bacon, Inc.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

CHAPTER 2

MEASUREMENT PROCEDURES STUDY

In order to measure rate of acquisition, the experimenter must first be able to measure acquisition. It will be recalled from Chapter 1 that acquisition in this thesis is defined as a change from the learner being unable to respond correctly to being able to respond correctly and continuing to do so on future occasions. One of the difficulties with this definition is defining "future occasions."

A commonly used procedure for measuring acquisition is the retention test administered at some point after the training period. However, there is little agreement as to what the retention interval should be. That is, there is little agreement as to how long the experimenter should wait before attempting to determine whether a learner has learned and remembered a response.

One of the first questions to be explored during the present experiments then was the question of how to determine whether or not a new response (that could be performed correctly during instruction) would also be performed correctly some time in the future. For the purposes of the present experiment, it was decided that a response would be deemed to have been added to the learner's repertoire (that is, acquired) if it could still be performed without prompting at least seven days after the final practice response. A seven day period was selected because previous research has shown that the recall of practised responses decreases rapidly over the 24-hour period following instruction and then decreases steadily over the next six days. Ebbinghaus (1885/1913) tested himself on lists of nonsense syllables and found recall decreased following practice by 42%, 66% and 75% after 20 minutes, 24 hours and 7 days respectively after instruction. When graphed, these results show "a smooth curve which falls rapidly at first and then more and more slowly" (Hunter, 1970, p. 129). This has been referred to as the curve of forgetting. Spitzer (1939) gave 3000 11-year old children eight minutes to study an article on bamboo plants. They were tested for immediate, 1-day and 7-day recall. He found that recall following the study period decreased to 53%, 38%, and 32% immediately, one day and seven days respectively following practice.

Because a response may be forgotten after it is performed correctly for the first time, a teacher may not classify it as acquired at that stage. To ensure that responses are learned and remembered in the classroom, some teachers require that a response be performed correctly on two or three occasions without assistance before classifying it as acquired (e.g., McLaughlin, Reiter, Mabee, & Byram, 1991; Schermerhorn & McLaughlin, 1997). This procedure is used so that the teacher can be sure that the response is learned and remembered rather than learned and forgotten.

It was decided for the present experiments that the retention interval between instruction and the retention test should be seven days. That is, acquisition was defined as a change from the learner being unable to respond correctly to being able to respond correctly without prompting seven days following instruction. The aim of this experiment is to determine whether a single test with a 24-hour retention interval was sufficient to produce a reliable measure of acquisition (learned and remembered for seven days), or whether a second and/or third test is required in order to establish that a response has been acquired and can be recalled seven days later.

METHOD

Learning Task

The learning task selected for each treatment was a set of 10 spelling words.

Setting

The Measurement Procedures Study was carried out in a Year 2 classroom in a Decile 7 school in Christchurch. Practice sessions were carried out at the participant's desks within their classroom. Individual presession testing and retention testing of each participant was carried out in a 2.5m by 4m office located within the school about 10m away from the participants' classroom. In the office was a desk facing the wall with two chairs next to each other. Each participant sat on the left chair with the experimenter sitting on the right chair.

Participants

A sample of 20 6- and 7-year old children was selected by drawing potential participants from the pool of normally developing 6- and 7-year old children in a single classroom. Participants ranged in age from 6 years and 3 months to 7 years and 4 months. Children were screened in the following manner. Those whose reading level was 12 months or more below their chronological age were excluded from the pool of potential participants on the grounds that the learning task would be too difficult for such students. Participants who were absent for more than one session during the study were also excluded. All participants in the class who met the inclusion criteria were included in the study.

Stimulus Materials

A pool of 30 unknown spelling words was identified for each of the 20 participants using the following procedure. Participants were seated at the table with the experimenter sitting to the right of them. They were presented with a pencil and recording sheet with their name written on it by the experimenter. The experimenter told the participant that he was going to teach them some new spelling words and needed to find some that they didn't know. The experimenter then told the participant the procedure. Words were selected from Levels 1 and 2 of *Learning to Spell* (Arvidson, 1960). The experimenter stated the target word, put it into a sentence, and then restated the target word. The participant then attempted to write the word on the test sheet. The experimenter repeated the word if the participant sought clarification. The experimenter did not provide feedback on the spelling response. This procedure was repeated for the next word. Praise was provided throughout the session for attending. The session ceased (a) when 30 unknown words were located, or (b) after approximately 15 minutes. If 30 unknown words were not located after 15 minutes the participant attended a second pre-experimental session the following school day.

Measurement Procedure

Each daily session began with the experimenter individually testing each participant on the words that he or she had practised 24 hours earlier. The experimenter stated the target word, said it in a sentence, and then repeated the target word. The participant then attempted to spell the word orally. The experimenter wrote down the letters that the participant said and marked the word as either correct or incorrect without the participant seeing. If the experimenter could not hear or understand what the participant said during the spelling of a word, the child was asked to start again. This procedure was repeated for each word to be tested. The retention test was completed after the participant had spelled (either correctly or incorrectly) all the words that had been practised 24 hours earlier.

Measurement Treatments

Three measurement treatments were investigated: correct 24-hours, correct 24- and 48-hours, and correct 24-, 48-, and 72-hours after instruction.

Measurement Treatment 1: Correct 24-hours after instruction. Each participant began each treatment with a practice session on 10 unknown words on a Monday. On Tuesday the participant was tested on the words practised on Monday. Words that were correct were withdrawn from the Tuesday practice set. The participant then practised the remaining words in Tuesday's practice session. (If all 10 words were correct on this or any subsequent test, practice ceased). On Wednesday the participant was tested on the words from Tuesday's practice session. Words that were correct were removed from the Wednesday practice set. The participant then practised the remaining words on the Wednesday practice session. On Thursday the participant was tested on the words from the Wednesday practice session. Words that were correct were removed from Thursday's practice session. The participant then practised the remaining words during the Thursday practice session. On Friday the participant was tested on the words practised in the Thursday session. In Treatment 1, participants received a 7-day retention test on the words that they spelled correctly on the 24-hour probe test 7 days previously. The procedure is shown in Table 1.

Table 1

General Procedure for Treatment 1

Week 1	Monday	Tuesday	Wednesday	Thursday	Friday
		Probe test of words practised 24 hours earlier	Probe test of words practised 24 hours earlier	Probe test of words practised 24 hours earlier	Probe test of words practised 24 hours earlier
	Practice session	Practice session	Practice session	Practice session	
Week 2	Monday	Tuesday	Wednesday	Thursday	Friday
		7-day retention test of words correct on Tuesday probe test	7-day retention test of words correct on Wednesday probe test	7-day retention test of words correct on Thursday probe test	7-day retention test of words correct on Friday probe test

Measurement Treatment 2: Correct 24- and 48-hours after instruction. The learner practised 10 spelling words over the course of one school week. These words were randomly selected from the participant's pool of 30 unknown words. During this treatment a word was removed from the practice set as soon as it was spelled correctly on the 24-hour probe test. The word was then tested a second time on the 48-hour probe test. If, on the 48-hour probe test, the word was spelled incorrectly it was returned to the practice set. If the word was spelled correctly on the 48-hour probe test it was tested again 7 days later.

Measurement Treatment 3: Correct 24-, 48- and 72-hours after instruction. The learner practised 10 spelling words over the course of one school week. These words were selected from the remaining words in the pool of unknown words. During this treatment a word was removed from the practice set once it has been spelled correctly on the 24-hour probe test. The word was then tested a second time 48 hours later. If incorrect it was returned to the practice set. If correct on the 48-hour probe test it was tested again the following day on the 72-hour probe test. If correct on the 72-hour probe test it was tested again 7 days later.

Experimental Design

The three testing treatments were administered in the following order: Measurement Treatment 2, Measurement Treatment 1, Measurement Treatment 3.

Practice Sessions

Spelling practice occurred during a daily, 20-minute whole class lesson. The children were taught by the experimenter to use the cover, copy, compare (CCC) method (Murphy, Hern, Williams, & McLaughlin, 1990) to learn their 10 spelling words. They were first taught to copy the word on a piece of paper from a written model, then to compare the spelling of the word they had written with the model to ensure it was written correctly, then to read the letters of the word they had written, and then to cover it with their non-writing hand and to write the word again without looking at the model. Following this, the child compared their response with the model to see if the response was correct. If it was, they moved on to the next word. If not, they were to repeat the procedure. Each participant was asked to practise their list three times.

The experimenter modelled and/or described this procedure to the participants at the beginning of each lesson while the participants were sitting on the mat. After this, each child was provided with a named piece of paper that described the spelling procedure and which listed the words to be practised that day. They were then instructed to return to their desks and to begin practising on the named blank piece of paper provided. If a child was unable to read the instructions or a spelling word, or required assistance, they were instructed to put their hand up and wait for the experimenter. Once participants had practised their set of words three times or after 20 minutes, whichever came first, they were instructed to hand both pieces of paper to the experimenter.

INTERSCORER AGREEMENT

Interscorer agreement was collected on 33% of the experimenter's written recordings of the participants' responses. Interscorer agreements were calculated on written recordings of 24-, 48-, and 72-hour probe tests and 7-day retention tests by totalling the number of agreements and dividing this by the number of agreements and disagreements and multiplying by 100. Interscorer agreement on the experimenter's written recordings of 24-, 48-, 72-hour and 7-day probe tests was 100%

RESULTS

Words Recalled

Table 2 and Figure 1 show that 97 words were correct on the 24-hour probe test for Measurement Treatment 1. Of these, 62 (64%) were retained in the 7-day retention test. Eighty two words were correct on the 24- and 48-hour probe tests for Measurement Treatment 2 and 67 (82%) of these were retained in the 7-day retention test. Seventy seven words were correct on the 24-, 48- and 72-hour probe tests for Measurement Treatment 3. Of these, 72 (94%) words were retained in the 7-day retention test. The number of words that were retained and forgotten on the 7-day retention tests under each of the measurement treatments for each participant is shown in Table 3.

Table 3 shows the number of words retained and forgotten on the 7-day retention test under each measurement treatment for each participant. The number of words retained ranged from zero words to nine words in Measurement Treatments 1 and 2, and from 0 words to 10 words in Measurement Treatment 3.

Table 2

Total Number of Words Practised and Meeting the Criterion, and Total Number and Percent of Words Recalled Under Each Performance Measurement Criterion

Measurement Treatment	Criterion	Total number of words practised	Number of words correct on 24-hour probe tests	Number of words correct on 48-hour probe tests	Number of words correct on 72-hour probe tests	Total number of words meeting the criterion	Number of words meeting the criterion recalled on 7-day retention test	Percent of words meeting the criterion recalled on 7-day retention test
1	24-hour probe test	200	97	-	-	97	62	64
2	24- and 48-hour probe test	200	96	82	-	82	67	82
3	24-, 48- and 72-hour probe test	200	93	83	77	77	72	94

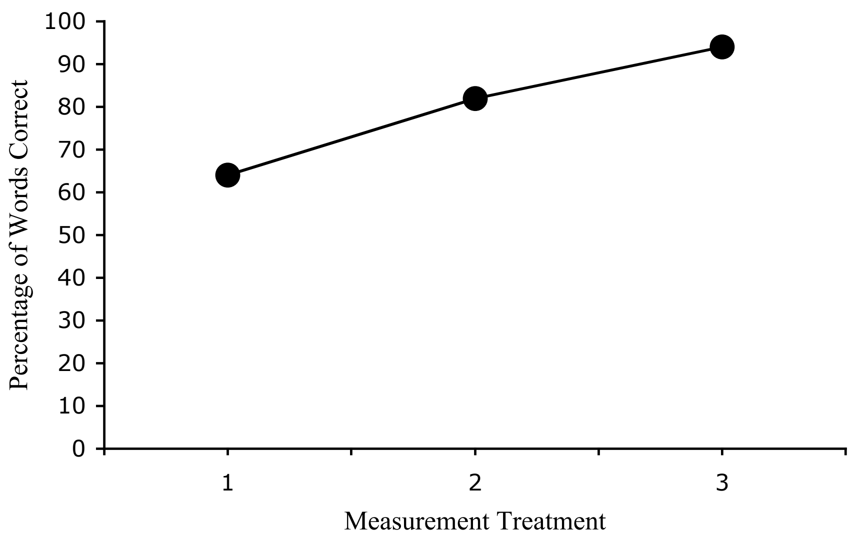


Figure 1. Percentage of words retained after seven days under each of the three measurement treatments.

Table 3

Number of Words Retained and Forgotten on 7-Day Retention Test per Performance Measure for Each Participant

	Number of words retained and forgotten on 7-day retention test								
	Measurement Treatment 1: 24-hours after instruction measure			Measurement Treatment 2: 24- and 48-hours after instruction measure			Measurement Treatment 3: 24-, 48- and 72-hours after instruction measure		
Name	Retained	Forgotten	Total	Retained	Forgotten	Total	Retained	Forgotten	Total
Mike	2	2	4	3	2	5	2	0	2
John	0	1	1	1	0	1	1	0	1
Mark	1	1	2	0	0	0	1	0	1
Johnny	1	0	1	1	0	1	0	0	0
Sam	3	1	4	5	1	6	1	0	1
Ant	4	1	5	3	1	4	3	0	3
Art	0	3	3	1	2	3	3	0	3
Ester	3	0	3	0	3	3	3	0	3
Cam	1	2	3	6	0	6	5	1	6
Sue	1	1	2	4	0	4	1	1	2
Nick	5	2	7	5	2	7	6	1	7
Troy	8	2	10	6	0	6	6	1	7
Mary	1	0	1	2	2	4	3	1	4
Jed	6	1	7	6	0	6	5	0	5
Taylor	0	0	0	1	1	2	1	0	1
Terry	4	3	7	5	0	5	5	0	5
Bart	0	2	2	1	0	1	1	0	1
Annie	9	0	9	9	0	9	10	0	10
Helen	6	4	10	2	0	2	7	0	7
Camden	4	4	8	5	1	6	5	0	5
Total	62	35	97	67	15	82	72	5	77

DISCUSSION

This experiment compared several procedures for determining when a new word had been acquired. It was found that 64% of the words that were correct on the 24-hour probe test were retained seven days after the 24-hour probe test, that 82% of words that were correct on both the 24- and 48-hour probe tests were retained seven days after the 48-hour probe test, and that 94% of the words correct on the 24-, 48- and 72-hours probe tests were retained seven days after the 72-hour probe test.

Forgetting followed the curve of forgetting (Hunter, 1970). Fourteen percent of learned and forgotten words were forgotten 24 hours after the probe test, and a further 22% of learned and forgotten words were forgotten over the next six days. That is, just over a third of the forgetting that occurred over seven days occurred within the first 24 hours.

There are two possible explanations for the differences in predictive accuracy between the measurement treatments. First, all words correct on the 24-hour probe test in Measurement Treatment 1 were tested for retention in the 7-day retention test. In Measurement Treatment 2 and 3 however, some of the words correct after 24 hours were incorrect on the 48-hour probe test in Measurement Treatments 2 and 3 or the 72-hour probe test in Measurement Treatment 3. These words were returned to the practice set and were not tested 7 days later because they had not been learned to criterion. The predictive accuracy of the 24- plus 48-hour probe test procedure and the 24-, 48- plus 72-hour probe test procedures therefore increased because the base number of words meeting criterion on the final test was progressively reduced prior to the administration of the 7-day retention test. Second, it is possible that the additional practice opportunities provided by the 48- and 72-hour probe tests aided retention although this effect is likely to have been small.

As Measurement Treatment 3 (24-, 48- and 72-hours after practice probe tests) provided the best prediction of retention (94%) of a word seven days after the final probe test, it was selected as the measure of acquisition for the first experiment.

REFERENCES

Arvidson, G. L. (1960). *Learning to spell: A manual for use with the N.Z.C.E.R. alphabetical spelling list*. Wellington, New Zealand: New Zealand Council for Educational Research.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University (Original work published 1885, translated 1913).

Hunter, I. M. (1970). *Memory*. London: Penguin Book Inc.

McLaughlin, T. F., Reiter, S. M., Mabee, W. S., & Byram, B. J. (1991). An analysis and replication of the Add-A-Word Spelling Program with mildly handicapped middle school students. *Journal of Behavioral Education, 1*, 413-426.

Murphy, J. F., Hern, C. L., Williams, R. C., & McLaughlin, T. F. (1990). The effects of CCC approach in increasing spelling accuracy with learning disabled students. *Contemporary Educational Psychology, 15*, 378-386.

Schermerhorn, P. K., & McLaughlin, T. F. (1997). Effects of the Add-A-Word spelling program on test accuracy, grades, and retention of spelling words with fifth and sixth grade regular education students. *Child and Family Behavior Therapy, 19*, 23-35.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*, 641-657.

CHAPTER 3

COMPUTER SET UP, PROGRAMMING AND GENERAL PROCEDURES

In Chapter 1 it was argued that teaching variables are often left uncontrolled in teaching experiments (Poling, Methot, & LeSage, 1995) and even when they are controlled they are often not recorded or reported. This is important because, if extraneous teaching variables are left uncontrolled, the experimenter cannot ascertain the effects of experimental teaching variables on rate of acquisition.

In addition, the measurement of instructional efficiency requires the experimenter to employ continuous measures of responding so the point when acquisition occurs can be detected. The safest way of ensuring procedural reliability is to record all instructional events and all learner responses during each teaching session. An experiment must therefore provide adequate levels of experimental control over teaching variables. To achieve the desired level of experimental control and procedural reliability over teaching variables in the present experiments it was decided to use appropriately programmed computers to control these variables (e.g., Experiments 3 and 4 in Chapters 7 and 8).

There are several advantages in using computers to administer and record instructional events. First, computers can provide a much greater level of experimental control over teaching variables than that which could be achieved by a human teacher (Gage & Berliner, 1998; Karsh & Repp, 1992). Second, a computer can also keep a much more detailed record of instructional events, learner responses and learner latencies than is possible using human observers. Third, employing computers to administer and measure the instructional events allows the experimenter to observe “additional interests” in what the participant is doing (Johnston & Pennypacker, 1993).

Given these advantages it was decided to administer all practice sessions using specially programmed Macintosh LC 575 computers for the experiments described in this thesis.

Computer Set Up

Macintosh LC 575 computers running System 7.6.1 with 12 MB of RAM, 256 bit colour screens and 80 MB hard drives were programmed for the present experiments using HyperCard 2.3 (Apple Computer, 1995). The two programmes used in the present experiments were part of a suite of six programmes written by Craig and Raewyn Saunders for the Learning in Young Children Project. All instructional programmes were in colour, animations were provided by calls to Macromind Director, and synthesized speech provided by calls to MacinTalkPro 2.

Spelling Words

The pool of words for the programme for 6-year old children (Spelling 6) consisted of 174 words from Level 1 of Elley, Croft and Cowie (1977). They were words with frequencies in children's writing of less than 85 but greater than 6 per 33 000 words. The pool of words for the more advanced Spelling 8 programme consisted of 200 words from Level 3 of the Alphabetical Spelling List Book 2 (1961). These words were selected by asking a sample of eight Year 4 teachers to identify the Level 3 words that Year 4 children might have difficulty in learning to spell.

Field Testing

Extensive testing and field testing was carried out to ensure the computer programmes worked in the manner intended for sustained periods of time. The main problem which was identified during this field testing was the fact that the presentation of more than about 90 Hypercard cards (records) in succession tended to exceed the available memory of the LC575 computers. This problem was managed by rebooting the computer at the start of each session.

Parameter Setting

The Spelling 6 and Spelling 8 programmes were extremely versatile. Instructional parameters could be set by the experimenter via a set of menu options on two administration cards prior to each session. These are shown in Figures 2 and 3. The following parameters could be altered without additional programming.

- Adding participants to and removing participants from the experiment
- Specification of the instructional mode; whether an acquisition lesson, a fluency-building lesson or a test
- Selection of the experimental design (whether alternating conditions or some other design)
- Adjustments to the MacinTalkPro 2 pronunciation of particular spelling words
- Selection of the MacinTalkPro 2 voice
- Selecting, from the total list of words, the specific word lists for a particular child in a particular experiment
- The duration of the lesson in either number of minutes or number of trials
- Whether or not a visual timer was displayed
- Adding or removing post-practice session games or movies
- Whether or not the HELP ME, SHOW ME, or SAY IT buttons were displayed (and therefore active)
- Whether or not an antecedent model (“Automatic Demo”), error-contingent model (“Automatic Correction”), correct answer feedback (✔), or error feedback (✘) were provided

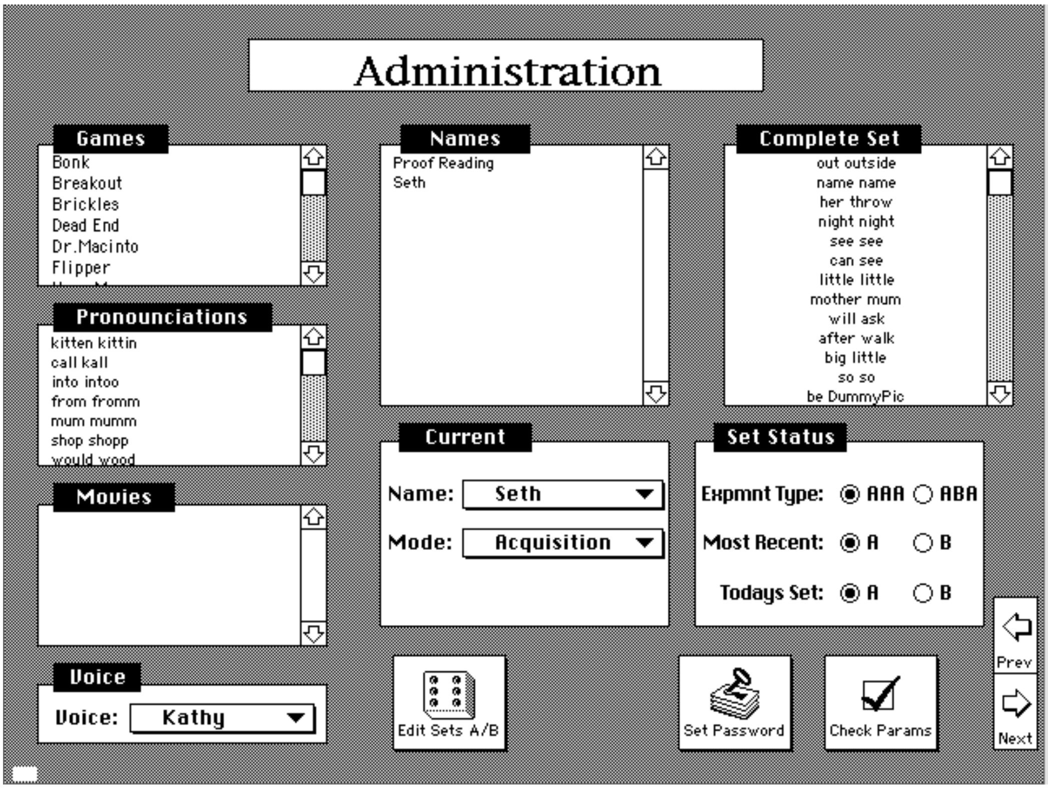


Figure 2. Administration screen 1 for Experiment 4 for Seth for the Spelling 6 Programme.

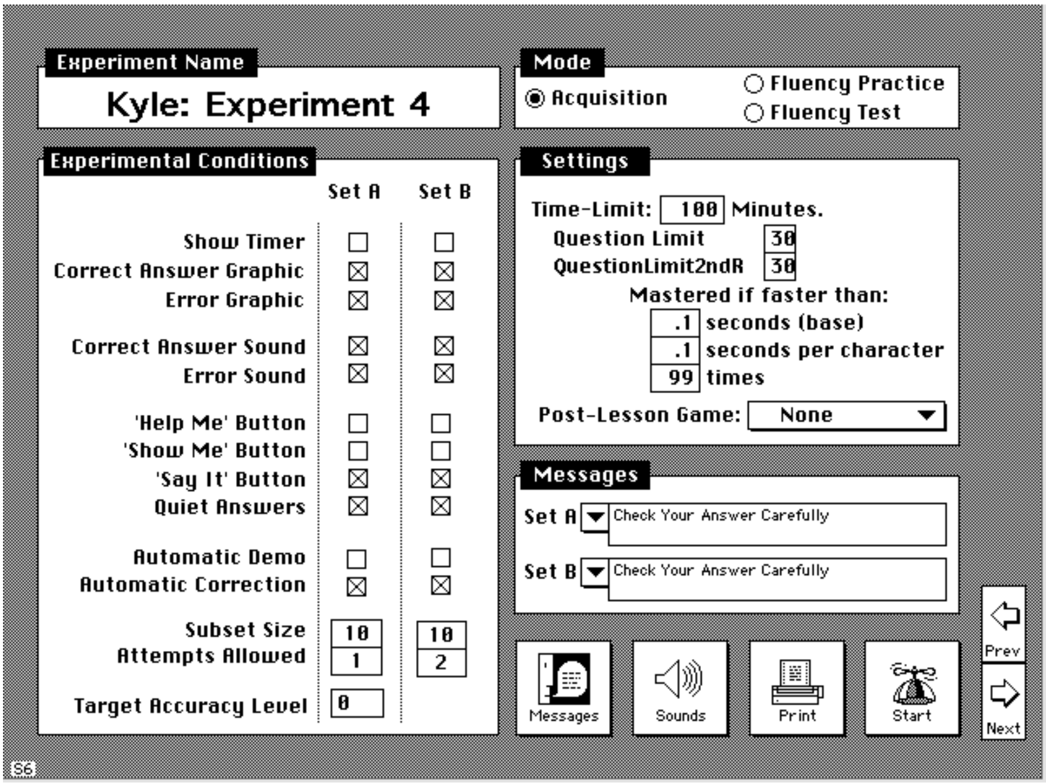


Figure 3. Administration screen 2 for Experiment 4 for Seth for the Spelling 6 Programme (Parameters in the “Settings” menu were not used and have been set to non-functional values).

- Whether or not auditory feedback was provided following correct and/or error responses
- Selection from a dictionary of sounds, the particular correct and error feedback sounds to be used in a particular experiment
- The Message Box statement
- The criterion for a word to be classified as mastered and removed from the child's list of practice words (e.g., number of times correct)

GENERAL PROCEDURE

General Procedure for all Experiments

Prior to the beginning of the investigations parameters were standardised for all experiments as shown in Figures 2 and 3. "Kathy" was selected as the computer-generated voice. "Acquisition" was selected as the instructional mode.

The "Correct Answer Graphic" (✔) and "Error Graphic" (✘) were activated. The "Correct Answer Sound" and "Error Sound" were also activated. The SOUNDS button opened the Sounds Lists Editor to manage the correct answer and error answer sounds. "Harp" was selected as the correct answer sound and "Ahem" was selected as the error sound.

The SAY IT button was activated. When the participant clicked this button the computer repeated the pronunciation of the target word. "Quiet Answers" was set so that the computer said the participant's spelling response followed by the pronunciation of the target word. "Subset Size" was set to 10. This was the number of words in the practice set for a session. "Question Limit" ended a session after a specified number of trials and was set to 30, that is, 10 words practised three times each. "Post-Lesson Game" was set at "None" as the experimenter managed this manually after the completion of each participant's spelling session.

General Procedure for Individual Experiments

The name of the experiment was entered in the “Experiment Name” as shown in Figure 3. Each child’s name was entered into the Spelling programme in the “Names” field on the administration card as shown in Figure 2. This (a) put the child’s name in a drop-down menu and (b) created a data file for that participant which recorded all instructional events that occurred during the course of each instructional session for that child. Prior to each child’s spelling session, that child’s name was selected from the “Current Name” drop-down menu on the Administration screen. To load the 10 words to be practised in the first session the EDIT SETS A/B button was clicked to open the Set Editor as shown in Figure 4. The Set Editor contained all the words from the pool of words in the column labelled “Spare.” It was possible to have two separate word sets for each participant. These were labelled “Set A” and/or “Set B.” Based on pre-experimental pretesting of participants, 10 words were added to either Set A or B for the child’s first session from the pool of words by selecting the word in the pool of words and clicking the appropriate ADD button. Once the 10 words to be practised by the participant for the first session were in either Set A or B the Set Editor was closed.

General Procedure Prior to Each Session

Prior to the start of each participant’s session the experimenter completed the following tasks. The correct spelling programme (Spelling 6 or Spelling 8) was opened and the child’s name was selected from the “Current Name” drop-down menu on the Administration screen. The words correctly spelled on that day’s probe test were removed and replaced with unknown words. Depending upon the experimental treatment, the number of responses per trial was entered into “Attempts Allowed” under “Experimental Conditions.”

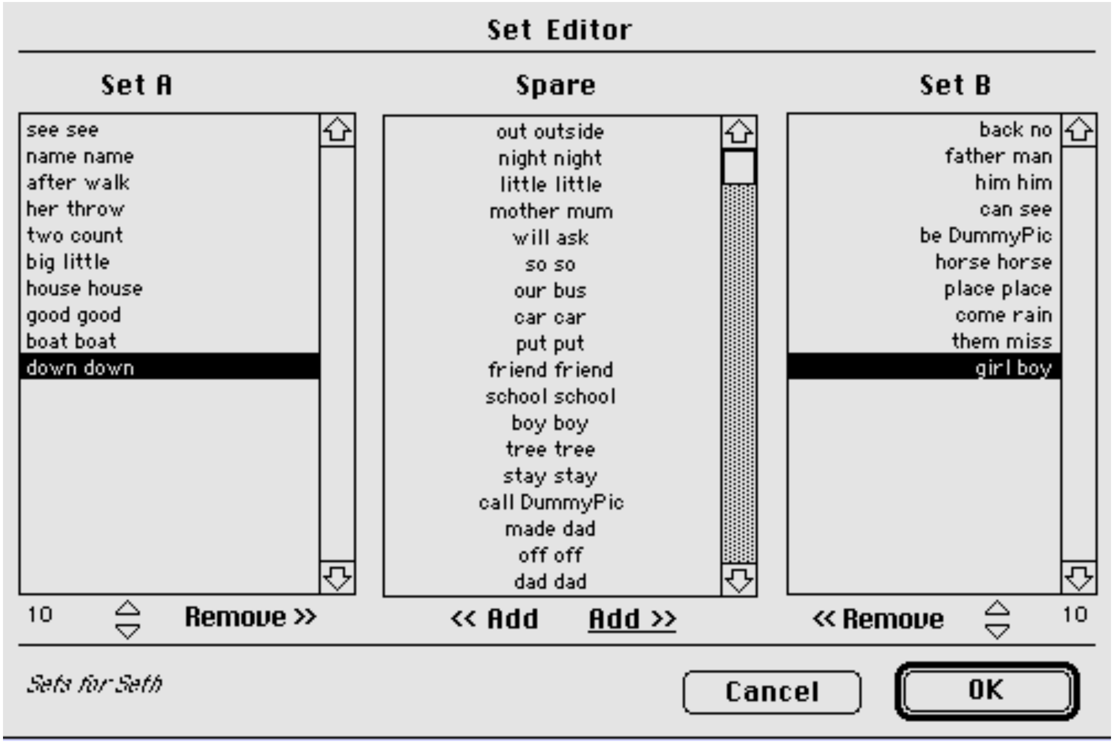


Figure 4. Set Editor screen for Experiment 4 for Seth for the Spelling 6 Programme. The first word in each line is the spelling word and the second word is the name of the picture which illustrates that word.

General Procedure During Each Session

Three children were run concurrently. The Spelling programme was set to the beginning-of-session hello screen for each participant by clicking the START button on the administration card as shown in Figure 3.

The participant sat at the computer terminal, put on the computer headphones and clicked the START button on the beginning-of-session Hello screen to begin the session as shown in Figure 5 (and Figure 11). The first word presented was the first word selected from the list of words to be learned that session.

At the beginning of each trial the computer presented a short sentence with one word missing, a copy of the missing word in the prompt box immediately above the sentence (if a model was programmed for that trial) and, in the case of Spelling-6,

Figure 5. S6 beginning-of-session "Hello" screen.

Figure 6. S6 practice screen (prior to responding).

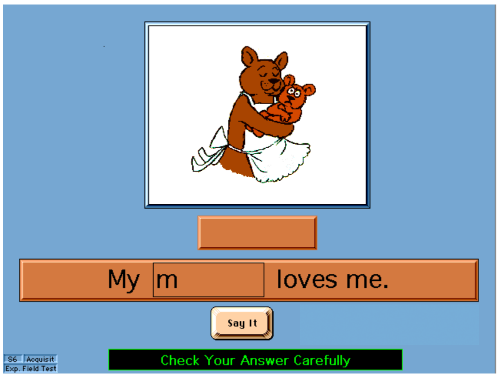


Figure 7. S6 first letter typed screen.

Figure 8. S6 completed response and correct feedback screen.

Figure 9. S6 completed response and error feedback screen.

Figure 10. S6 end-of-session "Good-bye" screen.

Figure 11. S8 beginning-of-session "Hello" screen.

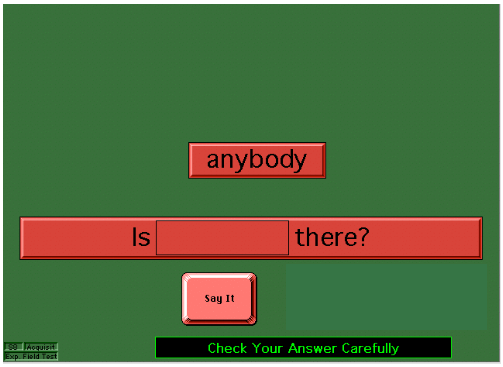


Figure 12. S8 practice screen (prior to responding).

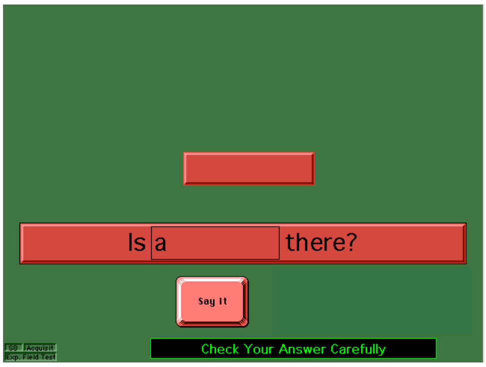


Figure 13. S8 first letter typed screen.

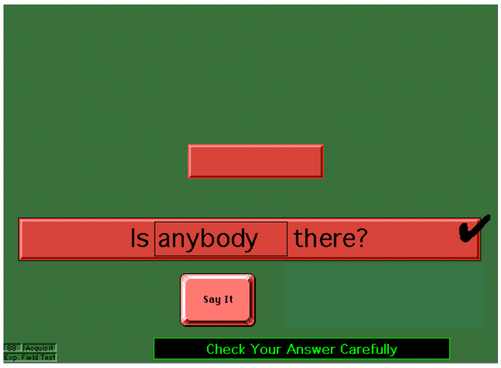


Figure 14. S8 completed response and correct feedback screen.

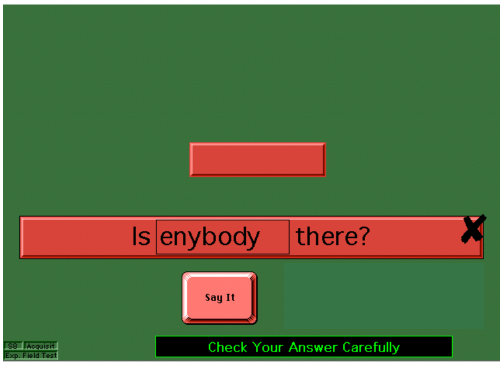


Figure 15. S8 completed response and error feedback screen.

Figure 16. S8 end-of-session "Good-bye" screen.

an illustration. These events are shown in Figures 6 and 12. The computer read aloud the complete sentence to the participant. The screen also contained a SAY IT button which, when clicked on, repeated the pronunciation of the target word.

As soon as the participant began to type their response any antecedent model on the screen disappeared as shown in Figures 7 and 13. As the participant typed the computer presented a typewriter-key sound and displayed the letter on the screen within the sentence. Corrections were possible using the DELETE key. Once the participant had completed their spelling of the word they pressed the RETURN key. A response concluded once they pressed the RETURN key.

In all experiments if the participant responded correctly the computer presented (a) feedback in the form of a tick (✔) at the end of the sentence as shown in Figures 8 and 14, and (b) the correct answer sound. A ✔ is the New Zealand convention for a correct answer. If the participant responded incorrectly the computer presented (a) feedback in the form of a cross (✘) at the end of the sentence as shown in Figures 9 and 15 and (b) the incorrect answer sound. The computer moved to the next screen (the next trial) after approximately one second if the participant responded correctly and after eight seconds if the participant responded incorrectly. In some experiments the child was asked to respond again following an error.

After the 10 words had been presented to the participant the computer re-presented the first word again and re-presented words two to ten in a new random order. The computer then re-presented the 10 trials a third time in a new random order. The spelling session was completed once the participant pressed the RETURN key after the thirtieth trial.

General Procedure Following Each Session

Immediately following the thirtieth trial the computer (a) said “No words left”, (b) presented the end-of-session Goodbye screen as shown in Figures 10 and 16, and (c) updated the participant’s data file. Figure 17 shows an annotated data-file printout for five trials from one session. Data collected included the participant’s name, date and start time of the session, the session mode, the current practice set, whether an antecedent model was presented, the target word, the participant’s responses (both primary and secondary), the time to respond (in seconds), whether the response was correct or incorrect, the number of responses allowed per trial, the number of responses per session, and the number of trials per session.

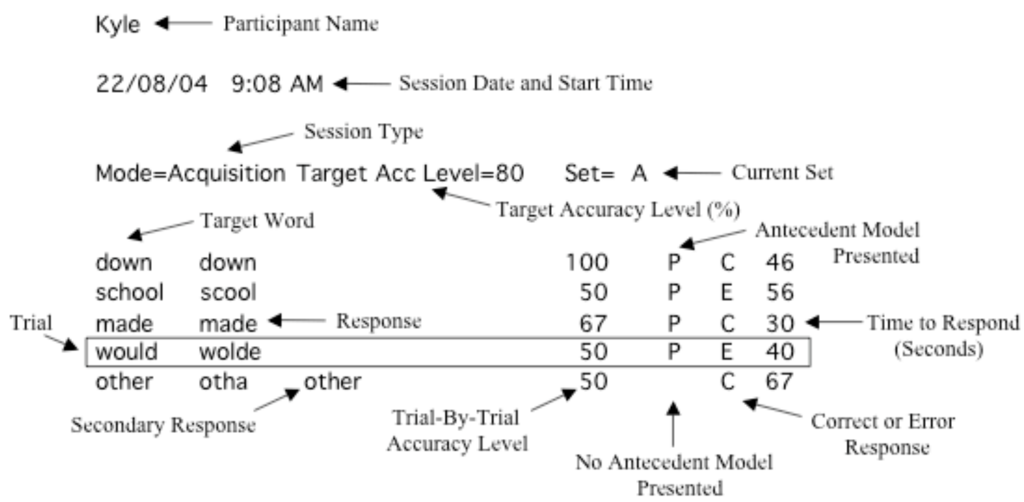


Figure 17. Five lines from a sample daily computer-generated end-of-session data file.

REFERENCES

Alphabetical spelling list: Book 2. (1961). Wellington, New Zealand: New Zealand Council for Educational Research.

Apple Computer. (1995). HyperCard (Version 2.3). California.

Elley, W., Croft, C. A., & Cowie, C. (1977). *A New Zealand basic word list*. Wellington: New Zealand Council for Educational Research.

Gage, N. L., & Berliner, D. C. (1998). *Educational psychology (6th ed.)*. Boston: Houghton Mifflin.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Karsh, K. G., & Repp, A. C. (1992). Computer-assisted instruction: Potential and reality. In N. N. Singh & I. L. Beale (Eds.), *Learning disabilities: Nature, theory and treatment*. New York: Springer-Verlag.

Poling, A., Methot, L. L., & LeSage, M. G. (1995). *Fundamentals of behavior analytic research*. New York: Plenum Press.

CHAPTER 4

A REVIEW OF THE RELATIONSHIP BETWEEN ACCURACY LEVEL DURING ACQUISITION AND RATE OF ACQUISITION

One of the variables that appears critical to rate of acquisition is feedback, particularly following errors (Kulhavy, 1977). However, in order to investigate the effects of feedback following errors it is necessary for the learner to produce incorrect responses during instruction. If there is a relationship between accuracy level during instruction and rate of acquisition, then allowing the accuracy level during instruction to vary will result in variations in the rate of acquisition. This also raises the question of how to control the level of correct and incorrect responding during instruction so that the learner makes a predictable proportion of errors during instruction.

One of the central debates in research on teaching concerns the amount of guidance that should be provided to the learner during the initial acquisition phase, that is, what should the proportion of errors during instruction be? It is possible to provide a high degree of guidance (scaffolding or prompting) during the initial acquisition phase. Under these conditions children make few errors (Day, 1987; Gast, Ault, Wolery, & Doyle, 1988; Miller & Test, 1989; Singleton, Schuster, Morse, & Collins, 1999; Taber & Glaser, 1962) and therefore few error corrections are required.

Behaviour analysts have tended to argue that errors should be avoided or at least reduced to low levels for three reasons. First, Skinner (1968) argued that instructional programmes should be constructed so that the learner is nearly always right. Skinner viewed learner errors from a punishment perspective and argued that it was the confirmation of corrects that was crucial to learning (Skinner, 1958). “Skinner (1954, 1958) has always insisted that it is the confirmation of the correct response which is crucial to learning” (Annett, 1969, p. 101). Attending to errors involves the presentation of corrections. These may be experienced by learners as aversive stimuli. If the accuracy level is low the learner "may become discouraged, motivation to practise may be reduced, and the pupil may begin to avoid the 'difficult' task" (Church, 1999, p. 26). Second, incorrect responses, once they begin to occur, tend to be repeated even when followed by corrective feedback. This is often referred to as error-perseveration (Grant & Evans, 1994). This is a particularly serious problem in the classroom where much practice is unsupervised. If incorrect responses are practised during unsupervised practice sessions, they may be strengthened to the point where additional instructional time is required to extinguish them. Third, practising incorrect responses uses up instructional time that could otherwise be spent practising correct responses.

Modern behaviour analysts, however, tend to argue that learners need to make some errors. Lindsley (1990) argued that some errors should be allowed so that learners do not become "addicted" to accuracy or perfection. He refers to errors as learning opportunities. If no errors are occurring then the teacher may be slowing the student down. If instruction is highly prompted, the teacher can't tell whether or not the learner could have moved more quickly through the instructional activities. "Pupils are likely to get to 'mastery' more quickly if allowed to make errors than if instructional conditions are arranged to avoid errors whenever possible" (O. White, personal communication, July 28, 2001). For example, a highly prompted teaching sequence might take eight steps to teach a particular spelling word whereas if the learner had been allowed to make errors he or she might have required only three or four responses with feedback to learn to spell the word.

Once errors do occur they need to be corrected otherwise the instructional stimulus may gain stimulus control over an incorrect response. There are many different ways of responding to errors: telling the learner they have made an error, telling the learner the correct response, asking the learner to try again (Grimes, 1981; Wolery, Bailey, & Sugai, 1988).

Highly-structured teaching procedures are variously referred to as instructivist, structured or errorless procedures. Examples include programmed instruction (Markle, 1991), Direct Instruction (Adams & Engelmann, 1996), Precision Teaching (Lindsley, 1990), the Personalized System of Instruction (Skinner, 1968) and Generative Instruction (Johnson & Layng, 1994).

Cognitive scientists argue that some errors should occur (thereby lowering the accuracy level) during instruction. Errors are seen as positive and a way of gaining insight into how learners are organizing their experiential world (Murphy, 1997). Hiebert and Carpenter (1992) argue that errors are a natural consequence of attempting to integrate new procedures with prior knowledge. Kulhavy (1977) argues that the main effect of feedback is to correct errors. According to Kulhavy errors interact with response certitude. Response certitude is the degree of confidence that the learner has that their response is correct. Kulhavy argues that feedback has its strongest effect on acquisition when response certitude is high but the response is incorrect.

There are many examples of teaching procedures with low levels of prompting. Examples include the various project methods of progressive education (Dewey, 1963), developmentally appropriate practice (N.A.E.Y.C., 1996), whole language (Goodman, 1994), discovery learning (Bruner, 1966), and constructivist teaching (Hiebert & Carpenter, 1992).

Whether a teacher provides a high or a low level of prompting during the acquisition phase has often been conceptualised as a dichotomy. However, it may be better viewed as a continuum from highly prompted responses within a structured instructional sequence towards lower levels of prompting within a less structured sequence. It is probable that most teachers fall somewhere near the middle of the continuum.

However, in the classroom it is often not possible to provide a high level of prompting during initial acquisition. In the classroom, a teacher often has to explain a new procedure to 30 or more children simultaneously. Under these conditions, the level of control over practice responses is likely to be low. During subsequent practice sessions, some children will have no difficulty in responding correctly while others will experience great difficulty responding correctly. Some children will make many error responses and will require frequent error corrections in order to acquire the skill or understanding which is being taught.

The debate over the degree of guidance that should be provided rests on the assumption that children should be practising correct responses during instruction. Presumably transfer of stimulus control can only occur when the correct response (in the presence of the stimulus) is reinforced. Here too there is a difference of opinion regarding the accuracy level that should be aimed for during instruction. Engelmann and Carnine (1991) argue that an accuracy level higher than 70% should be aimed for while in early reading instruction the goal is higher than 85%. In this thesis the proportion of correct responses during practice will be referred to as the accuracy level during instruction.

The debate over the degree of guidance rests on the assumption that rate of acquisition is influenced in part by the accuracy level during instruction. In order to examine the validity of this assumption it was decided to review experiments that (a) reported different accuracy levels during instruction in at least two treatments and (b) measured rate of acquisition under these treatments. It was considered important to find out whether the accuracy level during instruction is correlated with rate of acquisition because, if the two variables are correlated, then any experiment investigating other variables will need to control the accuracy level across experimental treatments.

AIM

The present review had two aims. The first was to determine whether a relationship exists between the accuracy level during instruction and rate of acquisition. The second was to gather ideas regarding ways in which the accuracy level during instruction might best be controlled in any experiment which attempts to measure the effects of different kinds of feedback during instruction.

METHOD

The literature search for the present review began with searches of the ERIC and PsycInfo databases using the keywords “errors and retention,” “errors and rate of learning,” “errorless and learning,” “error rate and learning,” “trial and error and errorless,” “programed instruct*,” “direct instruction and errors,” “accuracy,” “prompting,” “time delay,” “delayed prompting,” ”discrimination learn*” and “stimulus discrimination.”

Studies were included in the present review if they met the following criteria:

1. The behaviours taught were discrete responses.
2. There were at least two different treatments.
3. Rate of acquisition could be determined from the number of trials (or time) to criterion or from the level of recall or level of achievement following a fixed period of instruction.
4. The report included data that could be used to calculate the accuracy level of responding during instruction.

The tables of contents of the Journal of Applied Behavior Analysis, the Journal of Behavioral Education, the Journal of Learning Disabilities, and Research in Developmental Disabilities were searched for the last ten years to identify reports that might meet the inclusionary criteria above. The reference lists of each of the reports included in the review were also searched to identify further reports which might meet the criteria for inclusion.

Reports meeting these criteria were coded in terms of the number and age of participants, the learning task employed, independent variables, results, accuracy levels during instruction and rate of learning (or achievement level following a fixed period of learning).

Experiments (either between groups or within subjects) with more than two treatments were reported as pair-wise comparisons between the treatments with the highest and lowest accuracy levels during instruction. Experiments with multiple values of the independent variable were reported as pair-wise comparisons between the treatments with the highest and the lowest accuracy levels during instruction. Between-groups experiments and within-subject experiments were analysed separately.

Analysis of the Between-Groups Experiments

If the authors presented raw data on the number of errors and/or the number of correct responses during instruction (and included the total number of practice responses) for each treatment, then the error counts and/or the correct answer counts were converted to a percentage correct during instruction. This was done for each treatment.

If the authors presented data as percent of errors during instruction this was converted to percent correct by subtracting the percent of errors from 100 for each treatment. It is these mean treatment by treatment accuracy levels during instruction that are reported in Table 4. If the measure of learning was reported as number of errors, percent of errors, or number of responses correct on a posttest these were converted to percent correct for each of the treatment groups within the experiment.

Analysis of the Within-Subject Experiments

The data paths for each individual participant within each treatment within each experiment were examined to see if they conformed to a common pattern. That is, did Treatment A always produce a higher accuracy level during instruction than Treatment B?

This was accomplished by (a) calculating the mean accuracy level during instruction for each participant for each treatment within the experiment, and (b) comparing the accuracy levels to determine whether the treatment that produced the higher accuracy level during instruction did so for at least 90% of the participants. In all of the within-subject experiments the treatment comparisons meet this criterion except for Crosbie and Kelly (1994) where the difference between the two maximally different treatments was only six percent.

Because the data paths across individuals within experiments conformed to a common pattern in almost all experiments it was decided to aggregate scores across participants within treatments for each experiment. This was done in the following manner. First, the numerical value of each data point was identified for each participant within each treatment phase. Second, the value of the data points were summed for each participant, and divided by the number of data points in the phase to produce a mean within-treatment level for each participant. This was done for each treatment phase within each experiment. Third, raw data presented as the number of errors or the number of correct responses during instruction for each treatment were converted to a percentage correct during instruction. This was done for each participant within each treatment. Fourth, for each treatment phase within each experiment the accuracy level during that treatment was summed across participants and divided by the number of participants. It is these average within-treatment accuracy levels that are reported in Table 4 under the heading "Accuracy level during instruction."

If the author reported the average number of trials (or time) to criterion across participants for each treatment it is this figure that is reported in Table 4 under the heading of "Accuracy level on measure of learning." If the average number of trials (or time) to criterion was not provided by the author it was calculated by counting the number of trials (or time in minutes) to criterion for each participant for each treatment. These counts were aggregated across participants for each treatment and were converted to an average trials (or time) to criterion for that treatment. It is this figure that is reported in Table 4.

Correlation Between Accuracy Level During Instruction and Rate of Acquisition

Relationships between accuracy levels during instruction and accuracy levels on the measure of learning were investigated. Relationships were classified in the following manner. A positive relationship was said to have occurred if the treatment that produced the highest accuracy level during instruction also produced the highest accuracy level on the posttest or the least number of trials (or time) to criterion.

A negative relationship in an experiment was said to have occurred if the treatment that produced the lowest accuracy level during instruction produced the highest accuracy level on the posttest or required the least trials (or time) to criterion.

No relationship was said to have occurred if the difference between the mean accuracy level on the posttest following each treatment was five percent or less (or non significant).

RESULTS

The search described above located 42 studies (containing 57 experiments) that met the inclusionary criteria. It was not possible to identify a relationship in 18 of these experiments as the difference between the mean accuracy level during instruction of one treatment and another treatment was five percent or less. These experiments were removed from the analysis. Table 4 describes the basic procedures and the results of the 32 studies (39 experiments) where the mean accuracy level during the two instructional treatments differed by more than five percent. As can be seen from Table 4, 27 of the experiments used between-groups designs and 12 used within-subject designs.

Participants and Settings

Participants ranged from preschoolers to adults. Nine experiments classified the participants as adults, 28 as children, and 2 used both adults and children as participants. Twenty eight experiments classified the participants as normally developing and 11 as having some form of mental retardation or disability.

Twenty six of the experiments took place in a school setting, four in a university setting, and one in an office setting (Birnie-Selwyn & Guerin, 1997). Eight studies did not describe the experimental setting (Cheyne, 1966; Egeland, 1975; Ellis, Ludlow, & Walls, 1978; Gleason, Carnine, & Vala, 1991; Kaess & Zeaman, 1960; Krumboltz & Weisman, 1962; Kryzanowski & Carnine, 1980; Walsh, 1985).

Interobserver Agreement

None of the between-groups studies reported interobserver reliability data for either the responses made during instruction or the responses made on the posttest. Seven of the 12 time-series studies reported interobserver reliability however 5 did not (Birnie-Selwyn & Guerin, 1997; Crosbie & Kelly, 1994; Griffiths & Griffiths, 1976; Van Houten & Rolider, 1989, Experiments 1 and 3).

Learning Tasks

The learning tasks included principles of behaviour, psychological terms, rare French words, sign language, Kanji symbols, colour discrimination, geography, spelling (four experiments), naming and recognising alphabet letters and sounds (five experiments), reading words (two experiments), naming objects (two experiments), line form discrimination (two experiments), mathematics problems (eight experiments), and geometry (nine experiments).

.

Table 4
Basic Procedures and Results of Studies

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Bennett, Gast, Wolery, and Schuster (1986)	3 14- to 17-year olds with moderate to severe mental retardation	Manual sign of common objects	1: Percent correct 2: Trials to criterion	Multiple probe design	A: Time delay B: System of least prompts	Time delay more efficient	A: 99% B: 90%	Trials to criterion A: 180 B: 244	Positive
Birnie-Selwyn and Guerin (1997)	6 normally developing 4- to 7-year old children	Spelling	Percent correct	Within subjects	A: Critical difference B: Multiple difference	Critical difference more effective	A: 72% B: 57%	On posttest A: 87% B: 70%	Positive
Bradley-Johnson, Sunderman, and Johnson (1983)	39 preschool children	Naming letters of alphabet & numbers	Number of responses correct	Between groups	A: Delayed prompting B: Fading	Delayed prompting showed greater progress	A: 84% B: 64%	On 7-day posttest A: 99% B: 96%	None
Carnine (1976) Experiment 1	48 6-year olds (1st graders)	Recognising "e" & "i"	Number of responses correct	Between groups	A: Same practice set B: Separate practice set	Separated group made less errors	A: 34% B: 52%	On immediate posttest A: 21% B: 24%	None
Chalmers and Rosenbaum (1974)	100 under graduates	Recognising dimensions of shapes	Trials to criterion	Between groups	A: Reversal B: Non reversal C: Irrelevant	Reversal superior over non reversal	A: 31% B: 30% C: 48%	Trials to criterion A: 21 B: 48 C: 19	A-C Positive B-C Positive
Cheyne (1966) Experiment 1	80 under graduates	Geometry	Number of responses correct	Between groups	A: Fading B: No fading	Inconsistent results	A: 79% B: 25%	On immediate posttest A: 60% B: 52%	Positive

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Crosbie and Kelly (1994) Experiment 1	4 20-to 35-year old college students	Programmed text on behaviour analysis	Percent correct	Within subjects	A: No delay feedback B: Noncontingent delay feedback	Non contingent delay produced better performance	A: 68% B: 75%	Time to complete programme A:42 min B: 44 min	None
Duffy and Wishart (1987) Experiment 1	8 7-to 9-year old with Down's Syndrome	Shape discrimination	Number of responses correct	Between Groups	A: Fading B: Trial & error	Non handicapped children performed better	A: 96% B: 46%	On posttest A: 73% B: 66%	Positive
Experiment 2	8 normally developing preschoolers	Shape discrimination			A: Fading B: Trial & error	Overall, errorless of more value	A: 94% B: 55%	A: 93% B: 77%	Positive
Experiment 3	8 7- to 9-year old with Down's Syndrome	Nonsense figure discrimination			A: Fading B: Trial & error		A: 96% B: 46%	A: 93% B: 45%	Positive
Experiment 4	8 normally developing preschoolers	Nonsense figure discrimination			A: Fading B: Trial & error		A: 98% B: 64%	A: 93% B: 80%	Positive
Egeland (1975)	108 Preschoolers	Naming letters of alphabet	Number of letters correct	Between groups	A: Errorless + cue B: Errorless C: Trial & error	Errorless + cue more effective	A: 93% B: 80% C: 72%	On immediate posttest A: 91% B: 74% C: 76%	A-C Positive B-C None

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Elley (1966) Experiment 1	66 Teachers College students	Rare French words (rote task)	Number of responses correct	Between groups	Multiple Choice A: 4 options B: 2 options C: 1 option	Rote tasks: Efficiency reduced under high error rates	A: 43% B: 62% C: 100%	On immediate posttest A: 80% B: 82% C: 94%	A-C Positive
Experiment 2		Numerical series (meaningful task)			A: 4 options B: 2 options C: 1 option	Meaningful tasks: Initial error rates unrelated to test score	A: 72% B: 83% C: 100%	A: 74% B: 77% C: 67%	A-C Negative
Experiment 3	81 8- to 11-year olds	Task A: Difficult spelling words (rote task)	Number of responses correct		Multiple Choice A: 4 options B: 2 options C: 1 option		A: 36% B: 60% C: 93%	On 3-day posttest A: 34% B: 30% C: 41%	A-C Positive
Experiment 4		Task B: Age appropriate spelling words (meaningful task)			A: 4 options B: 2 options C: 1 option		A: 72% B: 83% C: 100%	A: 63% B: 67% C: 65%	A-C None
Ellis, Ludlow, and Walls (1978)	27 4th grade students	Kanji symbols	Percent errors	Between groups	A: Fading B: Trial & error	Fading took more trials to criterion	A: 98% B: 86%	No. of trials A: 36 B: 13	Negative
Everett (1977)	108 2nd grade children	Identifying line tilts and dots	Mean number of errors	Between groups	A: Errorless B: Trial & error	Errorless more effective	A: 91% B: 57%	Trials to criterion A: 32 B: 67	Positive
					C: Intra D: Inter		A: 65% B: 71%	A: 44 B: 55	Positive

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Gleason, Carnine, and Vala (1991))	95 elementary & middle school children	Geography - naming 7 countries	Number of responses correct	Between groups	A: Cumulative introduction B: Rapid introduction	Cumulative more effective	A: 93% B: 87%	On immediate posttest A:100% B: 98%	None
Godby, Gast, and Wolery (1987)	3 8- to 16-year olds with disabilities	Identifying common objects	1:Percent correct 2: Sessions to criterion 3: Time to criterion	Multiple probe design	A: Time delay B: System of least prompts	Time delay more effective	A: 96% B: 87%	Time to criterion A: 251 minutes B: 344 minutes	Positive
Griffiths and Griffiths (1976)	6 normal 5- to 6-year old children	Discriminating between "b,d" & "p,q"	1: Trials to criterion 2:Total errors	Within subjects counter balanced across tasks	A: Stimulus fading B: Trial & error	Fewer errors and trials to criterion with fading	A: 98% B: 61%	Trials to criterion A: 43 B: 59	Positive
Holcombe, Wolery, and Snyder (1994)	6 preschoolers with develop mental delays	Naming common objects on photographs	1: Percent Correct 2: Sessions to criterion	Adapted alternating treatments	A: CTD high fidelity B: CTD low fidelity	High fidelity more effective	A: 98% B: 89%	Time to criterion A: 55 minutes B: 83 minutes	Positive
Jacobs and Kulkarni (1966) Experiment 1	114 Junior high school children	Mathematics - solving equations	Number of correct responses	Between groups	A: Feedback B: No feedback	No feedback did not affect posttest	A: 93% B: 83%	On immediate posttest A: 32% B: 35%	None
Kaess and Zeaman (1960)	435 under graduates	Definitions of psychological terms	Number of errors	Between groups	Multiple choice A: 1 choice B: 2 choices C: 3 choices D: 4 choices E: 5 choices	Greater choice interferes with acquisition	A: 87% B: 70% C: 60% D: 51% E: 38%	On immediate posttest A: 90% B: 87% C: 87% D: 80% E: 77%	A-E Positive

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Krumboltz and Weisman (1962)	140 under graduates	Education statistics	Number of errors	Between groups	A: FR1 feedback B: FR67 feedback C: VR67 feedback D: FR33 feedback E: VR33 feedback F: No feedback	Higher schedules of feedback reduced errors during learning	A: 89% B: 89% C: 89% D: 87% E: 85% F: 81%	On immediate posttest A: 65% B: 65% C: 66% D: 62% E: 65% F: 62%	A-F None
Kryzanowski and Carnine (1980)	28 6-year olds (1st graders)	Recognising vowel sounds	Percent responses correct	Between groups	A: Massed practice B: Spaced practice	Spaced presentations superior	A: 74% B: 56%	On immediate posttest A: 38% B: 72%	Negative
Moore and Smith (1961) Experiment 1	62 13-year olds (6th graders)	Spelling	Number of words correct	Between groups	A: Immediate feedback B: Delayed feedback	No significant difference	A: Average 83% (on 4 weekly tests) B: Average 92% (on 4 weekly tests)	On 3-day posttest A: 82% B: 88%	Positive
Repp, Karsh, and Lenz (1990)	8 16- to 21-year olds with moderate to severe retardation	Naming numerical sets	1: Percent correct 2: Trials to criterion	Alternating treatments	A: Task demonstration B: Least-to-most prompting	Task demonstration superior	A: 86% B: 75%	Percent correct trials at 6-month follow-up A: 74% B: 58%	Positive
Repp, Karsh, Johnson et al. (1996) Experiment 2	10 10- to 19-year olds with moderate mental retardation	Reading sight words	1: Trials to criterion	Within subjects	A: One e.g. of S+, multiple e.g. of S- B: Multiple e.g. of S+ & S-	One e.g. of S+, multiple e.g. of S- superior	A: 69% B: 63%	Trials to criterion A: 32 B: 67	Positive

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Rosenstock, Moore, and Smith (1965)	92 6th grade children	Mathematics -set theory	1: Time to criterion 2: Percent of errors	Between groups	A: FR1 feedback B: FR5 feedback C: VR5 feedback D: No feedback	Schedules unrelated to achievement	A: 82% B: 61% C: 35% D: 34%	Time to complete programme A: 238 min B: 330 min C: 315 min D: 337 min	A-D Positive
Singleton et al. (1999)	4 adolescents with moderate mental retardation	Read grocery words	1: Percent correct 2: Sessions to criterion	Alternating treatments	A: Simultaneous prompting B: Antecedent prompting	Antecedent prompt more efficient	A: 68% B: 94%	Time to criterion A: 10 min B: 4 min	Positive
Smeets, Lancioni, and Striefel (1987)	4 9- to 13-year old children with handicaps	Mathematics problems	1: Percent correct 2: Time to criterion	Multiple baseline	A: Stimulus manipulation B: Delayed feedback	Stimulus manipulation favoured with generalisation	A: 99% B: 73%	Time to criterion A: 101 min B: 41 min	Negative
Smeets, Lancioni, and Streifel (1988) Experiment 1	16 preschoolers	Geometry	Trials to criterion	Between groups	A: Time delay (static cue) B: Time delay (dynamic cue)	No significant difference	A: 87% B: 97%	Trials to criterion A: 24 B: 17	Positive
Smeets, Lancioni, Streifel, and Curfs (1988) Experiment 3	40 4- to 5-year old children	Geometry	Trials to criterion	Between groups	A: Multi stimulus distinctive feature prompt B: Single stimulus distinctive feature prompt C: Single stimulus non distinctive feature prompt D: Single stimulus no feature prompt	Multi stimulus distinctive superior	A: 94% B: 85% C: 68% D: 72%	Trials to criterion A: 20 B: 29 C: Did not reach criterion after 30 trials D: Did not reach criterion after 30 trials	A-D Positive B-D Positive

Table 4
(Continued)

Authors and date	Participants	Task	Measure	Design	Independent variable	Results (reported)	Accuracy level during instruction (A)	Accuracy level on measure of learning (B)	Relationship between A and B
Strand (1989) Experiment 1	27 severely mentally handicapped 7- to 16-year olds	Line form discrimination	Trials to criterion	Between groups	A: S+ fading B: S- fading	No significant difference	A: 89% B: 82%	Trials to criterion A: 43 B: 70	Positive
Strand and Morris (1986)	16 7- to 14-year old children with severe mental handicaps	Identifying coloured shapes	Trials to criterion	Between groups	A: Graded prompting B: Stimulus fading C: Trial & error	Graded prompting and stimulus fading more effective	A: 100% B: 96% C: 65%	Trials to criterion A: 74 B: 90 C: 255	A-C Positive B-C Positive
Van Houten and Rolider (1989)) Experiment 1	10 6- to 7-year old children	Naming number facts	1: Percent correct 2: Sessions to criterion	Alternating treatments	A: Sequential presentation B: Rapid re-presentation	Rapid re-presentation more effective	A: 60% B: 70%	Sessions to criterion A: 10 B: 7	Positive
Experiment 3					A: Knee-to-knee B: Desk in between	Knee-to-knee more effective	A: 66% B: 55%	A: 8 B: 11	Positive
Walsh (1985)	21 22- to 59-year olds with severe mental retardation	Colour discrimination	Number of errors	Between groups	A: Fading B: Trial & error	Fading more effective	A: 95% B: 65%	Trials to criterion A: 43 B: 51	Positive

Independent Variables

The 39 experiments investigated the effects of a variety of treatments which had the effect of producing different accuracy levels during instruction. These are described in Table 4.

Experimental Results

As can be seen from Table 5, the 39 experiments generated 44 separate comparisons. Of these, 32 revealed positive relationships, 4 revealed negative relationships (Elley, 1966, Experiment 2; Ellis et al., 1978; Kryzanowski & Carnine, 1980; Smeets et al., 1987) and 8 were classified as indicating no relationship.

Table 5

Relationships Between Accuracy Level During Instruction and Measure of Learning

Experimental design	Positive relationship	Negative relationship	No relationship
Within subject	10	1	1
Between groups	22	3	7
Total	32	4	8

DISCUSSION

This literature review demonstrates that it is possible to have a positive relationship, a negative relationship, and no relationship between accuracy level during instruction and measure of learning, but that experiments reporting a positive relationship outnumbered the negative relationships by eight to one.

Cases Where There Was a Positive Relationship Between Accuracy Level During Instruction and Rate of Learning

The most common procedure employed to ensure a high accuracy level during instruction was the presentation of some form of prompt following the instructional stimulus.

The prompting procedures used included delayed prompting (Smeets, Lancioni, Streifel et al., 1988), stimulus fading (Griffiths & Griffiths, 1976), task demonstration (Repp et al., 1990) and antecedent prompt and test (Singleton et al., 1999).

Other procedures that produced high accuracy levels during instruction included concept examples (Chalmers & Rosenbaum, 1974), matching-to-sample training (Birnie-Selwyn & Guerin, 1997), high schedule of feedback (Rosenstock et al., 1965), rapid re-presentation of incorrectly answered questions (Van Houten & Rolider, 1989), position of the tutor (Van Houten & Rolider, 1989, Experiment 3), feedback immediacy (Moore & Smith, 1961) and multiple-choice treatments with varying numbers of distractors (Elley, 1966; Kaess & Zeaman, 1960).

In all of the cases where positive relationships were found, none of the stimulus control procedures that evoked the higher level of correct responding during instruction involved copying responses.

Cases Where There Was a Negative Relationship Between Accuracy Level During Instruction and Rate of Learning

A negative relationship between accuracy level during instruction and rate of acquisition indicates that lower accuracy levels during instruction were associated with higher levels of correct responding on the posttest. It is possible for participants to achieve lower levels of correct responding during instruction but higher rates of acquisition if the prompting procedure is inefficient. In other words, the participants may be provided with a greater number of prompts (which take up instructional time) than are actually necessary. This was observed in the Smeets et al. (1987) experiment where the high accuracy treatment produced an accuracy level during instruction of 99% but participants took, on average, 101 minutes to reach criterion whereas the low accuracy treatment had an accuracy level of 73% but participants took only 41 minutes to reach criterion. The high accuracy treatment required up to seven correct steps to complete a mathematics problem without the option of skipping steps whereas the low accuracy treatment provided the participants with feedback and only required them to correctly complete a problem.

The low accuracy treatment in Ellis et al. (1978) required participants to discriminate Kanji figures using a trial and error procedure. The accuracy level during instruction was 86% yet participants took only 13 trials to reach criterion. The high accuracy treatment, in contrast, required the participants to use a 26 step fading procedure and this resulted in a 98% accuracy level during instruction. However, this group took, on average, 36 trials to reach criterion. This suggests that participants may have become overdependent on the prompts by merely copying them. This was also observed in Elley (1966, Experiment 2) where participants copied the correct response thereby achieving almost errorless practice but obtaining only poor scores on the posttest. Jones and Eayrs (1992, p. 206) suggest that prompting "techniques may confine the person's attention to very narrow attributes of the stimulus associated with reinforcement." It seems likely that the participants in the Ellis and the Elley experiments may have focussed their attention primarily on the prompt rather than on the stimulus item *and* the prompt (as a supplementary stimulus). The Ellis and the Elley results support Kulhavy's (1977, p. 219) assertion that if "answers are easily accessible, students will spend their time copying them."

Kryzanowski and Carnine (1980) found that spaced practice for 6-year old participants produced lower accuracy levels during instruction yet higher posttest results when compared with massed practice. In the high accuracy (massed practice) treatment a stimulus item was presented to a participant and, if responded to incorrectly, error feedback was provided. The stimulus item was then re-presented to the learner. Under these conditions, it is highly likely that the error feedback functioned as a prompt for the second attempt. In this experiment, the accuracy level during massed practice was 74% and the posttest score was 38%. However, in the low accuracy treatment (where each particular item was presented in a spaced format) the feedback provided after a response functioned as a weaker prompt as other stimulus items were presented and responded to prior to the re-presentation of that particular stimulus item. Under these conditions the accuracy level during instruction was only 56% but the posttest score was 72%.

These results show that higher accuracy levels during instruction can be associated with lower rates of acquisition if the learner is (a) unable to skip unnecessary prompts or steps, (b) allowed to engage in copying responses, or (c) required to engage in massed practice with small sets of practice stimuli.

Cases Where There Was No Relationship Between Accuracy Level During Instruction and Rate of Learning

In eight of the comparisons, no relationship was observed between accuracy level during instruction and rate of learning. In most cases this was because the differences between the experimental posttest scores were not significant (e.g., Bradley-Johnson et al., 1983; Elley, 1966, Experiment 4). Egeland (1975) reported an eight percent difference in accuracy level during instruction between the errorless treatment and the trial and error treatment but found a non-significant difference in posttest performance. The two groups of 6-year old participants in Carnine (1976) achieved only 21% and 24% on the posttest even though the accuracy levels during instruction were 34% and 52%. This difference was not significant.

In the Gleason et al. (1991) experiment, participants in the cumulative introduction of items treatment achieved a 93% accuracy level during instruction while the participants in the rapid introduction treatment achieved 87%. Posttest results were 100% and 98%, respectively. However, participants in the cumulative introduction treatment had, on average, 102 responses with 8 errors while participants in the rapid introduction treatment engaged in

258 practice responses with 35 errors. So, while there was a six percent difference in accuracy levels during instruction, participants in the rapid introduction treatment required over two and half times the number of practice responses to achieve a similar posttest score.

Participants in the Crosbie and Kelly (1994) experiment achieved accuracy levels of 68% and 75% during the two instructional treatments yet spent similar amounts of time (42 and 44 minutes respectively) completing the programme. Once feedback was provided in the 68% treatment, the next item was presented. However, in the 75% accuracy level treatment, participants were provided with a 10 second delay between feedback and presentation of the next stimulus item. During this delay the computer screen displayed the question, the response and the feedback to that response. Participants report that during this 10 second interval their attention was spent reading the question, the response, and the correct response. Given that participants spent extra time studying the questions and answers, and potentially engaging in additional covert correct practice responses, it is not surprising that the accuracy level during instruction was higher during this treatment. However, the time taken to complete the programme in each treatment was similar. Any instructional efficiency gained by implementing a 10 second interval between items (which produced a higher accuracy level during instruction) may have been lost due to this interval taking up additional instructional time.

Two studies (Jacobs & Kulkarni, 1966; Krumboltz & Weisman, 1962) allowed the students to copy their responses. Both used programmed texts with the answers in the back of the text. Jacobs and Kulkarni state that participants were able to "look ahead at the correct answer before making their overt response." It is probably this behaviour (looking ahead and copying the answers to programme question) that accounts for these non-significant differences between the posttest scores between the control and the experimental treatments in both of these experiments.

Weaknesses of the Experiments

Experimental design. As 27 of the 39 experiments reviewed employed a between-groups design it is difficult to draw conclusions regarding the effects of different stimulus control procedures on either rate of learning or postexperimental achievement for the individual participants in any of these experiments.

Interobserver reliability. A weakness of all the between-groups and four of the time-series studies was the absence of interobserver agreement information. As Cooper, Heron and Heward (1987, p.91) note, it "is important to have interobserver agreement on the actual data from which conclusions will be drawn."

Level of stimulus control. None of the studies that manipulated accuracy level during instruction did so with the intent of producing a given level of accurate responding. This meant that studies involving a number of different kinds of experimental treatments had to be selected for this review. This has greatly complicated the attempt to determine whether there is a relationship between accuracy level during instruction and rate of acquisition.

Conclusion

This review has demonstrated that there is likely to be a relationship between accuracy level during instruction and rate of acquisition under certain instructional conditions. This seems to occur when higher accuracy levels during instruction are achieved by providing higher levels of non-copying prompts. The implication for the present series of experiments is that attempts to control the accuracy level in future experiments will need to employ some form of non-copying prompt.

There are, however, certain kinds of instructional arrangements which do not produce a positive relationship between the accuracy level during instruction and the measure of learning. These include conditions where (a) the learner is unable to skip unnecessary prompts and/or steps within a programme of instruction, (b) the learner is able to engage in copying responses, and (c) error corrections are followed immediately by re-presentation of the same question. This third condition however is probably just a variation of the second case where the participants engage in copying responses as they are able to keep the correct response in working memory following the correction.

REFERENCES

Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR*. Seattle: Educational Achievement Systems.

Annett, J. (1969). *Feedback and human behaviour*. Middlesex: Penguin Books Inc.

Bennett, D. L., Gast, D. L., Wolery, M., & Schuster, J. W. (1986). Time delay and system of least prompts: A comparison in teaching manual sign production. *Education and Training of the Mentally Retarded, 21*, 117-129.

Birnie-Selwyn, B., & Guerin, B. (1997). Teaching children to spell: Decreasing consonant cluster errors by eliminating selective stimulus control. *Journal of Applied Behavior Analysis, 30*, 69-91.

Bradley-Johnson, S., Sunderman, P., & Johnson, C. (1983). Comparison of delayed prompting and fading for teaching preschoolers easily confused letters & numbers. *Journal of School Psychology, 21*, 327-335.

Bruner, J. S. (1966). Some elements of discovery. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery: A critical appraisal*. Chicago: Rand McNally and Company.

Carnine, D. W. (1976). Similar sound separation and cumulative introduction in learning letter-sound correrspondences. *Journal of Educational Research, 69*, 368-372.

Chalmers, D. K., & Rosenbaum, M. E. (1974). Learning by observing versus learning by doing. *Journal of Educational Psychology, 66*, 216-224.

Cheyne, W. M. (1966). Vanishing cues in paired-associate learning. *British Journal of Psychology, 57*, 351-359.

Church, R. J. (1999). *Basic learning processes*: University of Canterbury, Department of Education.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus: Merrill Publishing Company.

Crosbie, J., & Kelly, G. (1994). Effects of imposed postfeedback delays in programmed instruction. *Journal of Applied Behavior Analysis, 27*, 483-491.

Day, H. M. (1987). Comparison of two prompting procedures to facilitate skill acquisition among severely mentally retarded adolescents. *American Journal of Mental Deficiency, 91*, 366-372.

Dewey, J. (1963). *Experience and education*. New York: Macmillan Company.

Duffy, L., & Wishart, J. G. (1987). A comparison of two procedures for teaching discrimination skills to Down's Syndrome and non-handicapped children. *British Journal of Psychology, 57*, 265-278.

Egeland, B. (1975). Effects of errorless training on teaching children to discriminate letters of the alphabet. *Journal of Applied Psychology, 60*, 533-536.

Elley, W. (1966). The role of errors in learning with feedback. *Journal of Educational Psychology, 36*, 296-300.

Ellis, W. D., Ludlow, B. L., & Walls, R. T. (1978). Learning, transfer, and retention of errorless fading versus trial-and-error teaching. *Psychological Reports, 43*, 553-554.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and application*. Eugene: ADI Press.

Everett, F. L. (1977). Role of errors in children's discrimination learning. *Perceptual and Motor Skills, 45*, 1187-1201.

Gast, D. L., Ault, M. J., Wolery, M., & Doyle, P. M. (1988). Comparison of constant time delay and the system of least prompts in teaching sight word reading to students with moderate retardation. *Education and Training in Mental Retardation, 23*, 117-128.

Gleason, M., Carnine, D. W., & Vala, N. (1991). Cumulative versus rapid introduction of new information. *Exceptional Children, 57*, 353-358.

Godby, S., Gast, D. L., & Wolery, M. (1987). A comparison of time delay and system of least prompts in teaching object identification. *Research in Developmental Disabilities, 8*, 283-306.

Goodman, K. S. (1994). Deconstructing the rhetoric of whole language. *Reading Research Quarterly, 4*, 330-332.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: HarperCollins.

Griffiths, K., & Griffiths, R. (1976). Errorless establishment of letter discriminations with a stimulus fading procedure in pre-school children. *Perceptual and Motor Skills, 42*, 387-396.

Grimes, L. (1981). Error analysis and error correction procedures. *Teaching Exceptional Children, 14*, 17-21.

Hiebert, E. H., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics* (pp. 65-93). New York: MacMillan Publishing Co.

Holcombe, A., Wolery, M., & Snyder, E. D. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4*, 49-73.

Jacobs, I. P., & Kulkarni, S. (1966). A test of some assumptions underlying programmed instruction. *Psychological Reports, 18*, 103-110.

Johnson, K. R., & Layng, T. V. (1994). The morningside model of generative instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction*. California: Brooks/Cole Publishing Company.

Jones, R. S. P., & Eayrs, C. B. (1992). The use of errorless learning procedures in teaching people with a learning disability: A critical review. *Mental Handicap Review, 5*, 204-212.

Kaess, W., & Zeaman, D. (1960). Positive and negative knowledge of results on a Pressey-type punchboard. *Journal of Experimental Psychology, 60*, 12-17.

Krumboltz, J. D., & Weisman, R. G. (1962). The effect of intermittent confirmation in programed instruction. *Journal of Educational Psychology, 53*, 250-253.

Kryzanowski, J., & Carnine, D. W. (1980). The effects of massed versus spaced formats in teaching sound-symbol correspondences to young children. *Journal of Reading Behavior, 12*, 225-229.

Kulhavy, R. (1977). Feedback in written instructions. *Review of Educational Research, 47*, 211-232.

Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children, Spring*, 10-15.

Markle, S. (1991). *Designs for instructional designers*. Champaign, IL: Stipes Publishing Company.

Miller, U. C., & Test, D. W. (1989). A comparison of constant time delay and most-to-least prompting in teaching laundry skills to students with moderate retardation. *Education and Training in Mental Retardation, 24*, 363-370.

Moore, J. W., & Smith, W. I. (1961). Knowledge of results in self-teaching spelling. *Psychological Reports, 9*, 717-726.

Murphy, E. L. (1997). *Constructivism: From philosophy to practice. Retrieved July 27, 2001*, from http://www.stemnet.nf.ca/~elmurphy/emurphy/cle.html

N.A.E.Y.C. (1996). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8: Position statement*. Retrieved 25 February, 2005, from http://naeyc.org/about/positions

Repp, A. C., Karsh, K. G., Johnson, J. W., & VanLaarhoven, T. (1996). A comparison of multiple versus single examples of the correct stimulus on task acquisition and generalization by persons with developmental disabilities. *Journal of Behavioral Education, 6*, 213-230.

Repp, A. C., Karsh, K. G., & Lenz, M. W. (1990). Discrimination training for persons with developmental disabilities: A comparison of the task demonstration model and the standard prompting heirarchy. *Journal of Applied Behavior Analysis, 23*, 43-52.

Rosenstock, E. H., Moore, J. W., & Smith, W. I. (1965). Effects of several schedules of knowledge of results on mathematics achievement. *Psychological Reports, 17*, 535-541.

Singleton, D. K., Schuster, J. W., Morse, T. E., & Collins, B. C. (1999). A comparison of antecedent prompt and test and simultaneous prompting procedures in teaching grocery words to adolescents with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 34*, 182-199.

Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969-977.

Skinner, B. F. (1968). *The technology of teaching*. New York: Meredith Corporation.

Smeets, P. M., Lancioni, G. E., & Streifel, S. (1988). Time-delay discrimination training with static and dynamic cues. *Journal of Experimental Child Psychology, 46*, 337-361.

Smeets, P. M., Lancioni, G. E., Streifel, S., & Curfs, P. M. G. (1988). Establishing a difficult discrimination through time delay: Some critical dimensions of prompts. *Journal of Experimental Child Psychology, 45*, 280-302.

Smeets, P. M., Lancioni, G. E., & Striefel, S. (1987). Stimulus manipulation versus delayed feedback for teaching missing minuend problems to difficult-to-teach students. *Research in Developmental Disabilities, 8*, 261-282.

Strand, S. C. (1989). S+ versus S- fading in teaching visual discriminations to severely mentally handicapped children. *Journal of Mental Defiency Research, 33*, 283-299.

Strand, S. C., & Morris, R. C. (1986). Programmed training of visual discriminations: A comparison of techniques. *Applied Research in Mental Retardation, 7*, 165-181.

Taber, J. I., & Glaser, R. (1962). An exploratory evaluation of a discriminative transfer learning program using literal prompts. *Journal of Educational Research, 55*, 508-512.

Van Houten, R., & Rolider, A. (1989). An analysis of several variables influencing the efficacy of flash card instruction. *Journal of Applied Behavior Analysis, 22*, 111-118.

Walsh, P. G. (1985). Teaching colour discrimination to the mentally handicapped. *Irish Journal of Psychology, vii*, 36-49.

Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional students*. Massachusetts: Allyn and Bacon, Inc.

CHAPTER 5

RELATIONSHIP BETWEEN ACCURACY LEVEL DURING INSTRUCTION AND RATE OF ACQUISITION

PART ONE: EXPERIMENT 1

It will be recalled from Chapter 4 that the learner must produce some errors during instruction if the learning researcher wants to investigate feedback following errors. Allowing the accuracy level during instruction to vary will result in variations in the rate of acquisition if there is a relationship between accuracy level during instruction and rate of acquisition. The experiments analysed in Chapter 4 suggest that there are many treatments where higher accuracy levels during instruction lead to faster rates of acquisition. The most common of these were prompt fading procedures of various kinds. These procedures ensured that stimulus control was transferred from the prompt to the spoken word by generating a relatively high proportion of correct non-copying responses during instruction and practice. This raises the question of how to control the accuracy level during instruction.

AIM

The aims of the present experiment were:

1. To measure the effects of manipulating the level of antecedent modelling (which varies the accuracy level during instruction) on rate of acquisition in order to ascertain whether the accuracy level during instruction will need to be controlled in subsequent experiments,
2. To ascertain whether it is possible to control the accuracy level during instruction by controlling the level of antecedent non-copying prompting.

METHOD

Participants and Setting

Nine Year 2 children participated in Experiment 1. Table 6 describes the characteristics of the participants.

Table 6

Description of the Characteristics of Participants in Experiment 1

Participant	Gender	Age (years, months)	Reading level[a]	Spelling programme
1 Carol	Female	6.6	12 (6.5 years)	Spelling 6
2 Scarlett	Female	6.7	11 (6.5 years)	Spelling 6
3 Sally	Female	6.3	6 (6.0 years)	Spelling 6
4 Mark	Male	6.5	13 (6.5 years)	Spelling 6
5 Sade	Female	6.8	18 (7.0-7.5 years)	Spelling 8
6 Gema	Female	6.0	13 (6.5 years)	Spelling 6
7 Becky	Female	6.3	8 (6.0 years)	Spelling 6
8 Tatum	Female	6.0	13 (6.5 years)	Spelling 6
9 Lee	Male	6.9	8 (6.0 years)	Spelling 6

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

Participants were drawn from a pool of normally developing Year 2 children in one of the two Year 2 classrooms in a Decile 7 urban-primary school. The school principal selected the Year 2 class. Participants were screened in the following manner. Children were excluded from the pool of potential participants if they (a) had a developmental delay of 12 months or more below their chronological age, or (b) had a reading level 12 months or more below their chronological age, or (c) were unable to print all the letters of the alphabet, or (d) were unable to achieve a fluency level of 20 letters per minute on the 26 alphabet keys on a computer keyboard after 15 minutes of instruction. Experimental participants were drawn at random from the pool of children who survived these screening tests.

Written informed consent was sought and obtained from the school principal, the classroom teacher, and the parents of potential participants. Oral informed consent was obtained from the participants. The identity of all participants was kept confidential by assigning an ID number and code name to each experimental participant. Experiments were conducted at three computer stations in the teachers' section of the school library.

Pre-Experimental Procedures

Pre-experimental testing. Each child was tested by the experimenter on a pool of words selected from the Learning in Young Children (LYC) Spelling 6 programme. Children were tested on the Spelling 6 word list in order of difficulty until 50 unknown words were found. If 50 unknown words were not located for a child, they were tested on the LYC Spelling 8 word list using the same testing procedure. Pre-experimental testing sessions ceased when either (a) 50 unknown words (either Spelling 6 or Spelling 8) were identified, or (b) after approximately 15 minutes. If 50 unknown words were not identified after 15 minutes, the child attended a second and, if necessary, a third pre-experimental testing session on subsequent school days.

Pre-experimental typing training. Prior to the experiment, participants spent three sessions (a) practising the location of the alphabetical keys, (b) learning how to operate the spelling programme, and (c) playing the computer games. During typing training a HyperCard letter-typing programme displayed a visual stimulus of a randomly selected letter of the alphabet as shown in Figure 18. The child pressed the corresponding letter key on the keyboard. If the child responded correctly the computer provided feedback in the form of an auditory stimulus and then presented the next random letter. If the child responded incorrectly the computer provided an auditory stimulus in the form of a computer generated voice saying "Try Again" and the letter stayed on the screen until the child responded correctly.

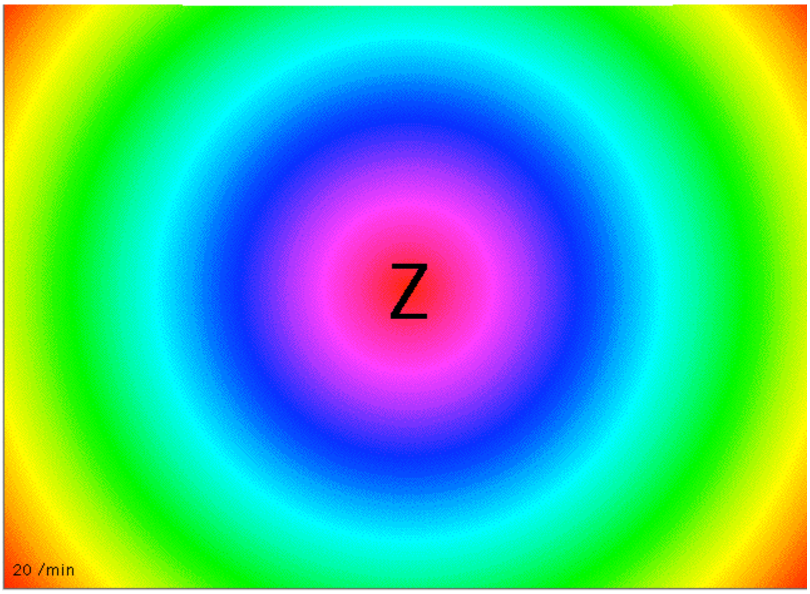


Figure 18. Screenshot of the HyperCard Typing Tutor developed for the present experiments.

During the session, a record of the mean response rate over the last three trials was inconspicuously displayed in the bottom left hand corner of the screen. Letter-typing training finished after the participant had reached an accuracy rate of at least 20 letters correct per minute for approximately half a minute. If an accuracy rate of 20 letters correct per minute was not achieved after five minutes, the child attended a second and, if necessary, a third letter-typing training session on subsequent school days.

Pre-experimental spelling training. The experimenter loaded five words into the spelling programme for the pre-experimental spelling training. If a participant was going to learn Spelling 6 words during the experiment, they practised Spelling 8 words in the pre-experimental training, and vice versa if a participant was going to learn Spelling 8 words. The participant sat at the computer as shown in Figure 19.

Figure 19. Participants working on the experimental spelling programme during Experiment 1.

The experimenter explained and modelled the procedure to be followed for the first word to be practised. The child was then shown the second word with the experimenter prompting when necessary. The training session was completed once the list of five words had been practised three times. If, after three practice sessions, the child was unable to complete the spelling programme independently, further sessions occurred until this requirement was met. After practising the spelling programme the child was taught, and given the opportunity to play, some of the simple computer games that could be played for five minutes at the completion of experimental sessions.

Measurement Procedures

Twenty four hours after each practice session (and prior to the next day's practice session) the experimenter individually tested each participant on the 10 words in their practice set. The experimenter stated the target word, put it into a sentence, and then restated the target word. The participant wrote the word on the test sheet. The experimenter repeated the word if the participant sought clarification. Feedback was not provided on the spelling response. This procedure was repeated for the next word. Praise was provided throughout the session for attending. The probe test included (a) the 10 words from the previous day's practice session and (b) any words that were to be tested a second or third time as part of the 48- or 72-hour probe test procedure. If the experimenter was unable to read any letter of the word that the participant had just written, he asked the participant to read it back to him. He then wrote each letter next to the participant's spelling word. This was done so that a reliability checker could read the participant's spelling of a word even if all the letters were not legible. The experimenter recorded whether the word was spelled correctly or incorrectly on a separate recording sheet. Words that were correct on this test were withdrawn from the practice session and replaced with unknown words prior to that day's practice session. Each word was then tested a second and third time in the following daily probe tests and, if correct on all three consecutive occasions, was classified as acquired.

General Procedure

Prior to the start of each school day the experimenter set up the computers and materials in the school library. The sequence of events for each participant is shown in Table 7. On the first day of a treatment the computers were set up ready for the participants to begin the spelling session. On subsequent days, as a probe test was to be administered, a new recording sheet (named and dated) and a sharp pencil was also organised for each participant. The order of participants was chosen by rotating the order so that different participants started first on different days. Depending on which participant was on the computer first, either the Spelling 6 or Spelling 8 programme was loaded. The participant's target accuracy level was set by changing the parameters on the participant's Administration Card (see Figure

Table 7

General Procedure for Each Treatment for Each of the Nine Participants

Week 1	Monday	Tuesday	Wednesday	Thursday	Friday
		Probe test of words practised 24 hours earlier	Probe test of words practised 24 (and 48) hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier
	Practice session	Practice session	Practice session	Practice session	Practice session
Week 2	Monday	Tuesday	Wednesday	Thursday	Friday
	Probe test of words practised 24 (and 48 and 72)[a] hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier	Probe test of words practised 24 (and 48 and 72) hours earlier
	Practice session	Practice session	Practice session	Practice session	

[a] The time difference between Friday and Monday was classified as 24 hours for the tests in Week 2.

3, Chapter 3). Just after 9.00 a.m. the experimenter collected the first three participants from the classroom and took them over to the school library.

Two participants chose and read a book when they arrived at the library while the experimenter administered the probe test for the first participant. After the final probe test of each treatment, participants were told which words they had acquired over that treatment. At the end of the daily probe test the experimenter (a) removed known words from the practice set on that participant's computer, and (b) replaced these words with unknown words from the pool of unknown words available from the administration menu. This kept the practice set to ten words each session. If a word that was spelled correctly on the 24-hour probe test was misspelled on the 48- or 72-hour probe test it was returned to the practice set. The experimenter then set the computer to the opening screen ready for the participant to begin.

The participant was seated at his or her computer and asked to start. This sequence of events was then repeated for the second and third participant in that session.

Computer programming modifications. For Experiment 1, the LYC spelling programmes were modified so that the computer provided an antecedent model of the word if the accuracy level (after each trial) was less than or equal to a predetermined target accuracy level. Depending upon the experimental treatment, a percentage number was entered into "Target Accuracy Level" field shown in Figure 3 (Chapter 3). The "Target Accuracy Level" controlled whether an antecedent model was presented or not for each trial. The computer provided an antecedent model of the target word if the accuracy level (after each trial) was less than or equal to a predetermined target accuracy level. For example, where the "Target Accuracy Level" was 80 percent the computer provided an antecedent model on the next trial if the actual accuracy level of correct responses was 80 percent or less across previously completed trials in that session.

If the actual accuracy level was less than or equal to the predetermined target accuracy level the computer presented (a) the sentence and a model of the word by displaying the word above the sentence (and below the picture for Spelling 6) as shown in Figure 20 and (b) the computer-generated voice (Kathy) saying the sentence.

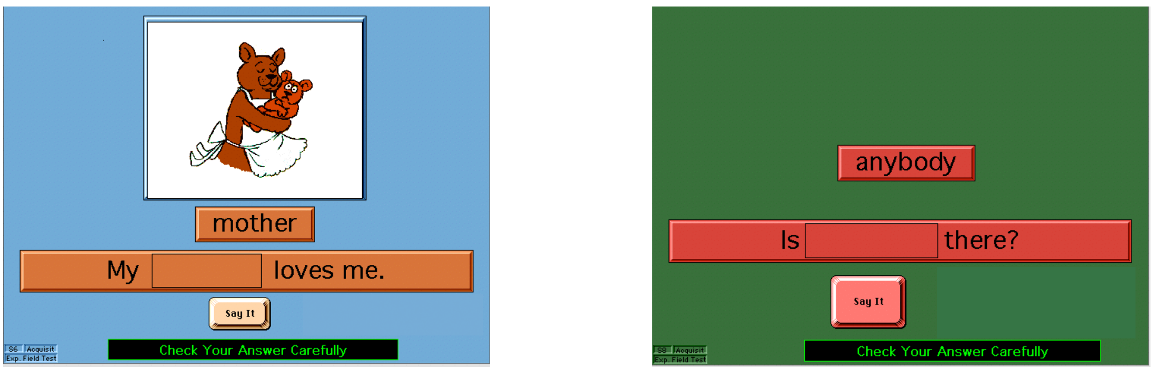


Figure 20. S6 and S8 trial screens with an antecedent model in Experiment 1.

If the actual accuracy level was greater than the predetermined target accuracy level the computer presented (a) the sentence as shown in Figure 21 and (b) the computer-generated voice (Kathy) saying the sentence. No antecedent model of the word was presented.

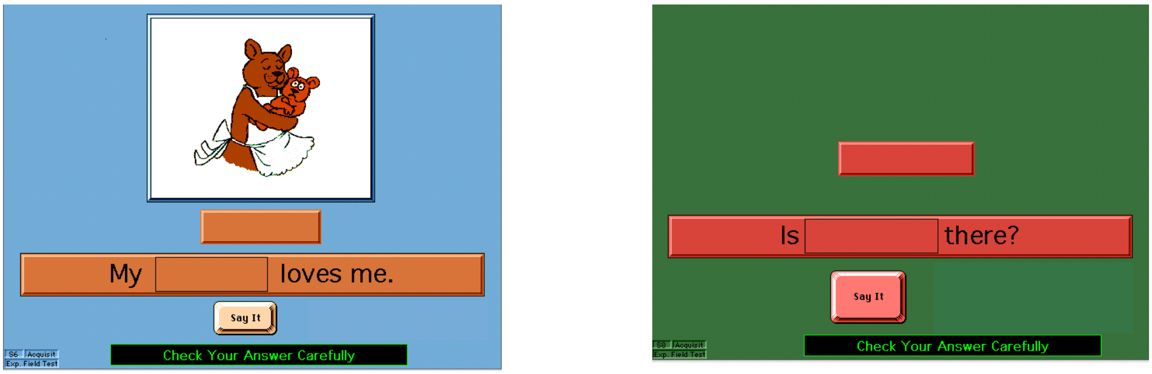


Figure 21. S6 and S8 trial screens without an antecedent model in Experiment 1.

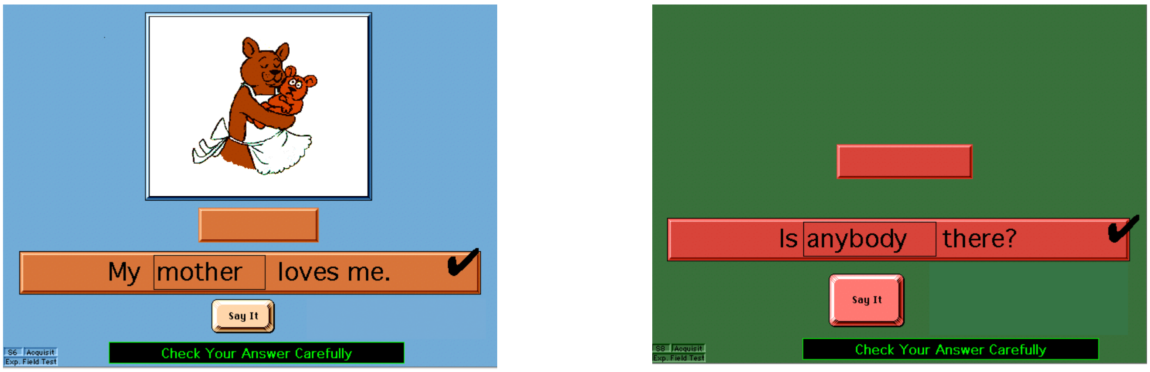


Figure 22. S6 and S8 completed response screens with correct feedback in Experiment 1.

Once the participant began to type their response the computer removed the model of the correct spelling from the screen. The participant completed their response by pressing RETURN. If the participant responded correctly the computer provided (a) feedback in the form of a ✔ at the end of the sentence as shown in Figure 22 and (b) the correct answer sound. If the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at the end of the sentence and a model of the word above the sentence as shown in Figure 23 and (b) the incorrect answer sound.

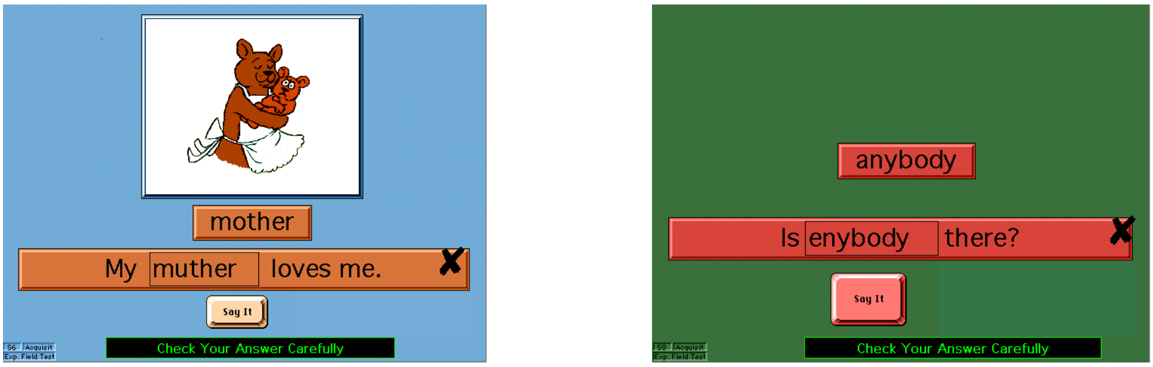


Figure 23. S6 and S8 completed response, error feedback and error-contingent model screen in Experiment 1.

After the 10 words had been presented to the participant the computer re-presented the first word again and re-presented words two to ten in a random order. The computer then re-presented the list of words a third time using this same procedure. The spelling session was completed once the participant pressed RETURN after the thirtieth trial. Once a participant had completed their spelling session the experimenter changed the computer application so the participant could play a computer game. After five minutes the experimenter told the participant that it was time to go back to the classroom and asked them to send over the next participant.

In order to evaluate the social validity of the experiments the experimenter interviewed each participant at the end of each treatment. To determine whether the participant enjoyed the experiment the experimenter asked, "When you were learning your spelling on the computer over the last few days did you, a) not enjoy it at all, b) enjoy it a bit, or c) enjoy it a lot?" To determine how difficult the participant found each experimental treatment the experimenter asked each participant, "When you were learning your spelling over the last few days, did you find the work on the computer easy, middle or hard?"

Experimental Design

A three-phase counter-balanced multiple-sequence across participants design was used. This design is an adaptation of the multiple-baseline multiple-sequence design (Noell, Gresham, & Gansle, 2002). The experiment consisted of three treatments: "65% target accuracy level during instruction," "80% target accuracy level during instruction," and "95% target accuracy level during instruction" replicated across nine learners. The order in which each participant experienced each treatment is shown in Table 8. Each treatment lasted until each participant had acquired 10 spelling words. All participants were run individually.

Table 8

Order of Treatments for Each Participant in Experiment 1

	Order of treatments		
Name	65% Target Accuracy Level	80% Target Accuracy Level	95% Target Accuracy Level
Sally	1st	2nd	3rd
Gema	2nd	3rd	1st
Carol	3rd	1st	2nd
Becky	1st	2nd	3rd
Scarlett	2nd	3rd	1st
Tatum	3rd	1st	2nd
Sade	1st	2nd	3rd
Lee	2nd	3rd	1st
Mark	3rd	1st	2nd

RESULTS

Interscorer Agreement

A second-year teacher trainee conducted the accuracy checks on the scoring of 25% of participants' responses on the probe tests. Interscorer agreement was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying this by 100. The range of interscorer agreement was from

94% to 100% with a mean of 99%.

Procedural Reliability and Treatment Integrity

A second-year teacher trainee conducted a procedural reliability check on 25% of the sessions to ensure that the participant received the correct level of modelling. Procedural reliability was assessed on agreements and disagreements between the within-session target accuracy level printout and the experimenter's recording sheet of a session within each treatment. Procedural reliability was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying by 100. Procedural reliability was 100%. Treatment integrity was assessed by viewing the computer printouts showing the participant's target accuracy level per treatment and their actual accuracy level over that treatment. The target and actual accuracy level of each treatment for each participant can be seen in Table 9. Actual accuracy levels across all treatments ranged from 21% (Sade) to 86% (Becky). The mean actual accuracy levels in the 65%, 80%, and 95% target accuracy level treatments were 66%, 66%, and 70% respectively, and there were wide variations between the target accuracy level and the actual accuracy level of individual children. The mean actual accuracy level treatment across all treatments was 67%. Given this lack of experimental control over the actual accuracy level during instruction, it was concluded that treatment integrity was not achieved.

Results

Six of the participants experienced all three treatments. Scarlett and Mark experienced two treatments and Lee one. Scarlett and Mark were absent from school for several days and ran out of time to complete all treatments. As Lee took 10 sessions to acquire eight words under the 80% target accuracy level treatment it was decided that the spelling words were too difficult. Lee was withdrawn from the experiment following his 10th session.

Rate of acquisition. Figure 24 shows the cumulative number of words correct on the

Table 9

Actual Accuracy Level and Mean Trials to Criterion for Each Participant Under Each Treatment

	Treatment A		Treatment B		Treatment C	
	65% Target Accuracy Level		80% Target Accuracy Level		95% Target Accuracy Level	
Participant	Actual accuracy level	Mean trials to criterion	Actual accuracy level	Mean trials to criterion	Actual accuracy level	Mean trials to criterion
Sally	72	7.5	83	6.0	78	5.4
Gema	63	4.4	75	3.2	82	5.5
Becky	62	6.0	78	4.2	86	4.4
Carol	58	5.2	58	5.5	54	5.4
Tatum	72	6.4	68	6.7	81	5.1
Sade	49	4.3	25	4.9	21	5.5
Scarlett	83	5.2	- [a]	-	72	6.0
Mark	-	-	72	9.0	83	6.8
Mean	66	5.6	66	5.6	70	5.5
SD	(11.1)	(1.1)	(19.6)	(1.9)	(22.1)	(0.7)
Lee	-	-		-	48	13.5

[a] Participant did not participate in that treatment.

24-, 48- and 72-hour probe tests for each of the participants under each treatment. The number of words acquired (the number of words correct on the 72-hour probe test) over all treatments ranged from 8 to 16 for Spelling 6. Sade practised Spelling 8 words and acquired between 11 and 12 words within each treatment. Excluding Lee, the participants acquired a mean of 11.3 words per treatment. Lee acquired eight words after 10 school days, at which time the treatment was terminated.

The between-phase data paths for each participant were almost parallel in most cases. This is to be expected given that treatment integrity was not achieved and there were few differences in accuracy level during instruction from one phase to the next.

Instructional efficiency. It will be recalled from Chapter 1 that instructional efficiency was defined as the number of trials to criterion. Trials to criterion was calculated by counting the number of practice trials on a response required in order to acquire that response. Mean trials to criterion was calculated for each participant for each treatment by dividing the number of practice trials on the acquired words in a treatment by the number of acquired words in that treatment. Table 9 shows the mean number of trials to criterion for each participant. Mean trials to criterion across treatments and participants ranged from 3.2 (Treatment B for Gema) through to 9.0 (Treatment B for Mark). The mean trials to criterion across participants were 5.6, 5.6, and 5.5 in Treatments A, B and C, respectively. The mean trials to criterion across all participants across all treatments was 5.6 (SD 1.3). There was little difference between the accuracy levels and rates of acquisition across treatments.

Effects of the model. In order to determine the cause of the failure to achieve treatment integrity a close inspection was made of the computer-generated printouts for each participant. These revealed that the model did not control the correct response to the degree expected. The results of the analysis are presented in Table 10 which shows the percentage of responses correct following the presentation of a model for the first round of a session (the first presentation of each word in a session) and for all three rounds in a session. The level of stimulus control exercised by the model ranged from 36% for Sade to 84% for Scarlett.

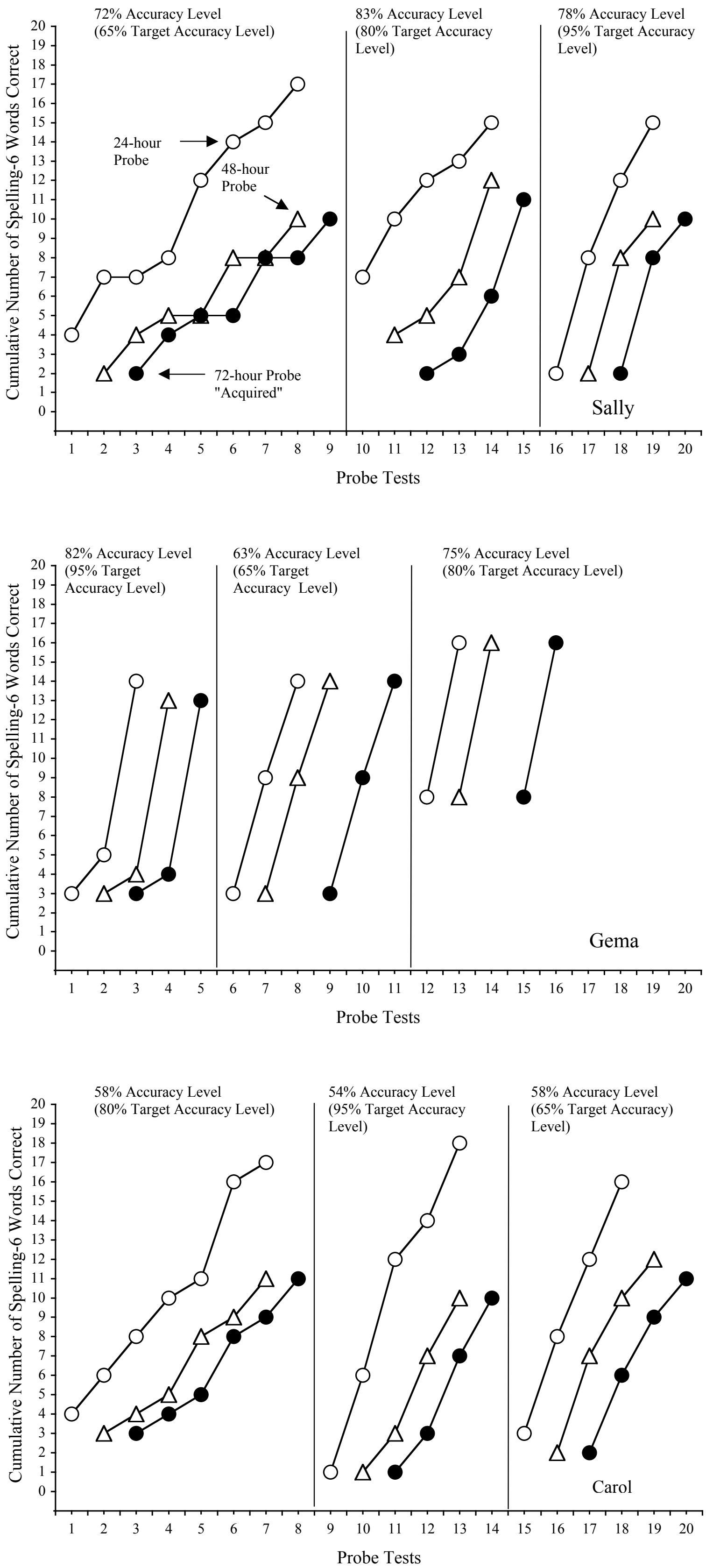


Figure 24. Cumulative number of words correct on the 24-, 48-, and 72-hour probe tests for each of the participants under each treatment in Experiment 1.

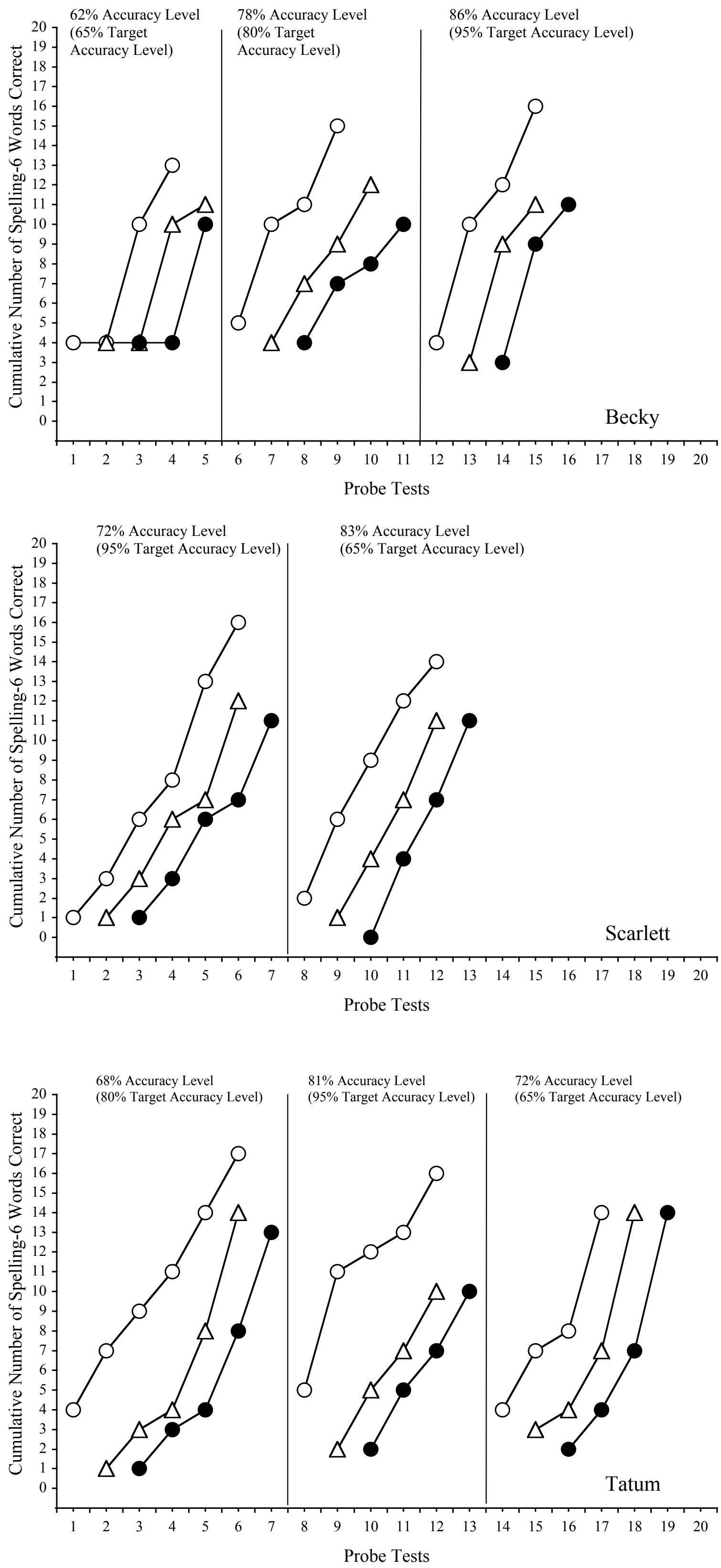
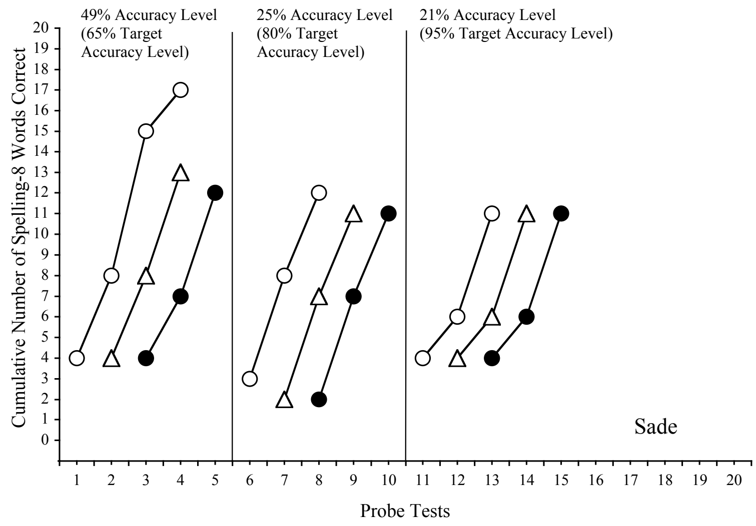
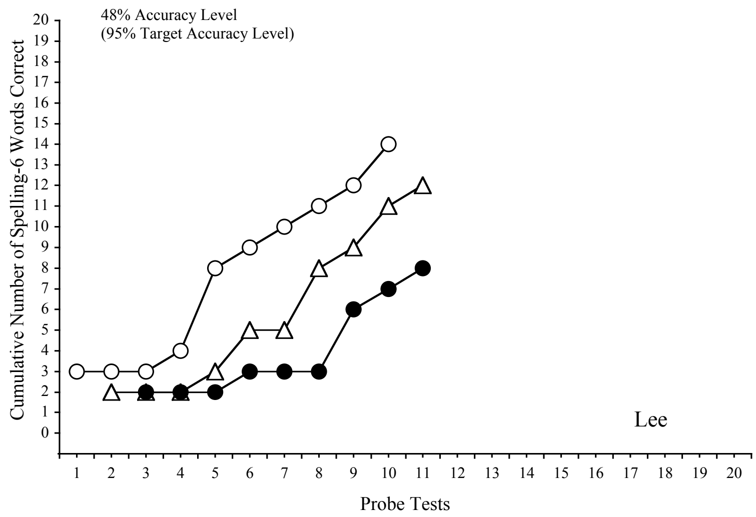


Figure 24. (Continued).





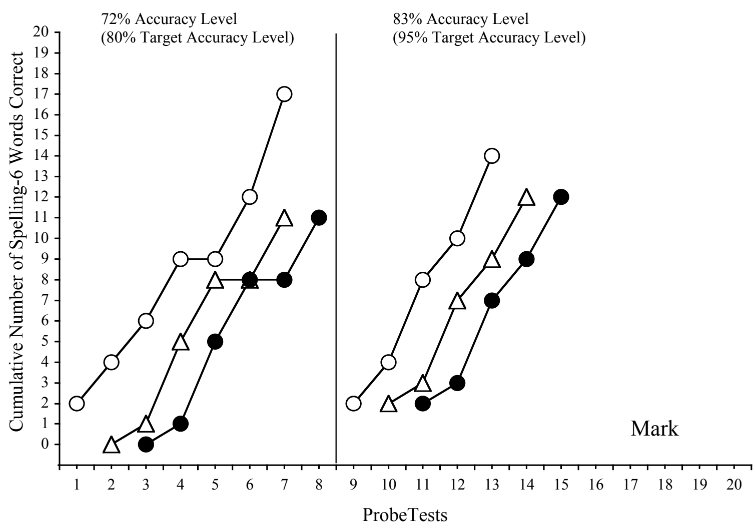


Figure 24. (Continued).

In only three cases (Sally, Becky, Scarlett) was the model effective more than 80% of the time. The mean percentage of correct responses following the model for all presentations was 68%.

Table 10

The Mean Percentage of Responses Correct Following a Model During the First Presentation of Word in a Session, and During all Presentations of a Word in a Session

Name	Mean percent of responses correct following model			
	During first presentation			During all presentations
	Acquired words	Non-acquired words	All words	All words
Sally	66	81	73	81
Gema	79	58	64	75
Carol	62	48	57	57
Becky	90	85	89	82
Scarlett	75	76	75	84
Tatum	66	42	57	74
Sade	0	41	33	36
Mark	56	26	42	51
Mean	62	57	61	68
Lee	83	40	44	50

DISCUSSION

Treatment Integrity Not Achieved

In this experiment treatment integrity was not achieved. Although the computer-programming algorithm worked because the model was presented as intended, experimental control over the accuracy level during instruction was not achieved. There appear to be two main reasons why treatment integrity was not achieved. The first of these relate to the modelling procedure employed and the second to an uncontrolled variable that operated during the experiment.

Non-copying model. It was assumed prior to the experiments that the model of the word in the antecedent position would already have stimulus control over correct responding. This was not the case. Participants only responded correctly 68% of the time, on average, after being presented with the model in the antecedent position. Failure to respond correctly following the model often occurred even when attending behaviour was high. This is almost certainly due to the fact that the model was a non-copying model. As soon as the participant began typing, the model disappeared from the screen. In a copying response the visual stimulus of each letter remains in view while the participant responds. This type of responding was not possible with the model disappearing from view once the participant pressed a key. In this experiment, some of the participants used mnemonic behaviour in order to remember letter sequences but others had not acquired this skill. Mnemonic behaviour is "overt or covert activity that produces supplementary stimuli which facilitate remembering" (Donahoe & Palmer, 1994, p. 359). After the model of the word had disappeared from the screen some participants were often observed whispering the letters of the word to themselves prior to pressing the letters on the keyboard. Sometimes they were successful in saying all the letters correctly and sometimes successful in saying only some of the letters correctly. Oftentimes a participant could be heard to say the letters correctly prior to responding only to say them incorrectly after beginning to type the word. Spelling a word correctly using a mnemonic strategy is considerably more difficult than copying especially for 6- and 7-year olds. If many of the children were unable to remember the correct letter sequence for many of the words throughout the period while they were typing their answers this would likely explain why the model was not always effective in prompting the correct response.

Experimental confound. Analysis of the lesson printouts further revealed that the experiment contained an uncontrolled extraneous variable. The model of the correct spelling of the spoken word was presented not only as an antecedent prompt but also as a consequence in the form of an error-contingent correction. This created a confound. For words that were acquired, transfer of stimulus control could occur in either of two ways. First, transfer from the antecedent model (as was originally planned) or, second, from the error-contingent model. This appears to have resulted in variability across the participants with respect to what they attended to most closely. For example, some participants on some occasions simply typed some vaguely relevant letters and then attended closely to the error-contingent model to determine the correct spelling. These apparently "careless" errors greatly reduced the accuracy level for some children in some experimental treatments. It also resulted in the transfer of control from the antecedent model on some words, and transfer from the error-contingent model on other words. This can be seen from the lesson printouts where misspellings on an early round in a lesson were sometimes followed by an unmodelled correct spelling on the next round.

In addition to the lack of control over prompting, an analysis of the data from Experiment 1 revealed two further poorly controlled variables.

Task difficulty. The first 50 unknown words tested from the Spelling 6 or the Spelling 8 Programme were selected as the learning task for each participant. The first 30 of these were randomly assigned to three sets (one for each treatment) with 10 words in each set. However, random assignment was not sufficient to control for word difficulty within each treatment. Because word sets were not matched for difficulty in terms of either the number of letter pairs in each word or the number of letter pairs incorrect on each word, some word sets were harder than others for a particular participant. For Gema, the first 10 words in Treatment A consisted of 56 letter pairs of which 30 letter pairs were incorrect on the pretest, Treatment

B consisted of 55 letter pairs with 33 letter pairs incorrect on the pretest, and Treatment C consisted of 59 letter pairs with 17 letter pairs incorrect on the pretest. It can be seen that the Treatment C set had approximately half as many letter pairs to acquire than the other two sets. As a result of this, Gema only required 3.2 trials to criterion on Treatment C words but required 5.5 trials to criterion on Treatment B words.

In some cases, individual words were probably too difficult for individual participants because there were too many letters following withdrawal of the prompt for the participant to recall. Case (1978, p. 440) states that young children "are incapable of dealing with very many items of information at one time." The actual number of letters within the word to be acquired ranged from one to eight in the Spelling 6 programme and one to nine in the Spelling 8 word list. It seems highly likely that some of the words consisted of a greater number of letters than these 6- to 7-year old children could "hold in mind" while typing.

The words in Lee's set were clearly too difficult for him. Treatment A words for Lee contained a total of 53 letter pairs with 32 letter pairs unknown on the pretest. That is, 60% of letter pairs were unknown. This is a mean of 3.2 letter pairs per word. This compares with Tatum's Treatment A words which had a total of 55 letter pairs and only 22 letter pairs incorrect (40% letter pairs unknown or a mean of 2.2 letter pairs per word). It seems that for Lee, many of the words in his list simply contained a greater number of letters than he could remember while attempting to type these words.

Finally, some words were inherently more difficult than others. For example, regardless of the treatment, four out of five participants who practised the word *other* did not acquire it. Scarlett was the only participant to acquire *other* and it took her 12 trials when her mean number of trials to criterion for all words was only 5.6.

Attending. The fact that participants sometimes responded incorrectly, even though they were presented with a model, may have been partially due to variability in the participants' attending behaviour. Donahoe and Palmer (1994, p. 153) define attending as "the failure of environmental stimuli to guide behavior even when they are apparently adequate to do so." Variability in the participant's attention was observed on a number of occasions during the course of Experiment 1.

In part, this was due to the session length of approximately 15-20 minutes. This was too long for some of the 6-year old participants. By the third presentation of a word some participants began to state that they were getting tired and/or bored. This was in part due to distractions while working on the spelling programme.

Prior to the experiment attempts were made to control attending responses by providing each participant with headphones to reduce extraneous stimuli. Despite this, it is clear that interference from other stimuli affected attention. Participants were observed attending to other participants and their computer screens. This interference was also observed when a class came into the library and a visiting child saw one of the participants and called out their name. The participant immediately ceased working on his/her spelling programme, swivelled the chair in the direction of the child and began interacting with them.

During the present experiment participants were seated on swivelling office chairs. Swivelling on the chair made it easy for the participant to turn and face other participants and computer screens. It was also a novel response that appears to have been reinforcing.

PART TWO: EXPERIMENT 2

Experiment 1 failed to answer the question of whether there is a relationship between accuracy level during instruction and rate of acquisition because of methodological weaknesses. These weaknesses included an uncontrolled extraneous variable, a lack of experimental control over the attending behaviour of participants, and poor control over the difficulty level of the learning task. It was decided therefore to make a number of methodological changes and repeat the experiment.

Experiment 1 failed throughout to sufficiently control the accuracy level during instruction. This was partially due to the fact that participants were presented with a model of the correct spelling in both the antecedent and consequent position. In Experiment 2 it was decided to remove the error-contingent model. If the participant responded incorrectly, they would be presented with feedback in the form of a cross (✘) and the incorrect answer sound. No model of the correct spelling of the word would be presented to the participant once they had responded incorrectly. Instead they would be presented with the next trial.

The attending responses of children in Experiment 1 were not managed very well because they were free to attend to other participants and their computer screens. Participants sat next to one another on swivel chairs and could see participants and their computer screens from where they sat. It would have been preferable to move the computer stations so they were not next to one another. Unfortunately, due to the layout of the school library this was not possible. It was therefore decided to place a screen between each computer station. It was hoped this would prevent participants from being able to see other participants and their computer screens. It was also decided to replace the swivel chairs with fixed chairs to prevent participants from swinging and spinning on their chairs.

To better control the difficulty level of the words in each treatment, it was decided to remove known difficult-to-learn words such as *other* in Spelling 6 from the pool of unknown words. In addition, words that were very similar (e.g., *would* and *could*), if randomly assigned to the same set, were re-assigned to two separate sets following the rule that easily confused items should be placed in separate practice sets (Engelmann & Carnine, 1991).

The probe testing procedure in Experiment 1 tested all words that were practised 24 hours earlier as well as words correct 48 and 72 hours earlier. A disadvantage of this procedure was that participants often faced a daily test of over 20 words. This took approximately 12-15 minutes. Participants were then to practice their spelling words on the computer for approximately 20 minutes. This meant the entire session took approximately 30-35 minutes. In Experiment 2 it was decided to measure acquisition on only the first 10 words practised on the first session to decrease the length of time required to administer the probe tests. This meant that the probe test would measure acquisition of a maximum of 10 words in any one test.

In Experiment 1 the accuracy level during instruction was calculated on all words practised and all words practised were measured for acquisition. This allowed any relationship between the accuracy level during instruction (calculated on all words practised) and all words acquired to be identified. In Experiment 2 it was decided to report the relationship between accuracy level during instruction and rate of acquisition in two ways. The first was to report the relationship between the accuracy level during instruction calculated on (a) all words practised and (b) those words acquired out of the first 10 words practised. This report was used because it might identify a relationship between the accuracy level during instruction and words acquired. However, while the accuracy level was calculated on all words practised, not all words were measured for acquisition. This was a weakness as it was likely that some non-measured words were acquired, and these non-measured words could effect the relationship. However, as they were not measured, the effect could not be identified. A solution to this was to employ a second relationship measure. This reported the relationship between the accuracy level during instruction calculated on only the first 10 words practised and those words acquired out of the first 10 words practised. This measure allowed any relationship to be identified.

AIM

The aims of Experiment 2 were as follows:

1. To ascertain whether it is possible to control the accuracy level during instruction by removing error-contingent models,
2. To ascertain whether it is possible to gain tighter experimental control over (a) the attending behaviour of participants and (b) the difficulty level of the practice words,
3. To measure the effects of manipulating the level of antecedent modelling (which varies the accuracy level during instruction) on the rate of acquisition in order to ascertain whether the accuracy level during instruction will need to be controlled in further experiments.

METHOD

Participants and Setting

As in Experiment 1, nine Year 2 children participated in the experiments. These children were selected from the other Year 2 class within the same Decile 7 school that participants from Experiment 1 had been selected. Children in this class were, on average, approximately 10 months older. Table 11 describes participant characteristics. Selection and screening of participants was the same as in Experiment 1.

One participant was withdrawn from the experiment after two sessions and one probe test. During the third computer spelling session, Alice started crying and would not attempt the target word on the computer. She stated that she didn't want to spell the target word incorrectly and that Carlos (the participant sitting next to her) had already finished which meant that she was slow. The session was terminated at that point. The experimenter subsequently spoke with both her classroom teacher and her father. They both stated that she

Table 11

Description of the Characteristics of Participants in Experiment 2

Participant	Gender	Age (years, months)	Reading level[a]	Spelling programme
1 Adwin	Male	7.4	21 (8.0 years)	Spelling 8
2 Kate	Female	7.2	23 (8.5-9.0 years)	Spelling 8
3 Carlos	Male	7.1	26 (9.5-10.0 years)	Spelling 8
4 Alice	Female	7.1	26 (9.5-10.0 years)	Spelling 8
5 Patrick	Male	7.2	22 (8.0-8.5 years)	Spelling 8
6 Tania	Female	7.4	26 (9.5-10.0 years)	Spelling 8
7 Dean	Male	7.6	26 (9.5-10.0 years)	Spelling 8
8 Nicky	Female	7.4	23 (8.5-9.0 years)	Spelling 8
9 Casey	Male	7.2	22 (8.0-8.5 years)	Spelling 8
10 Bernard	Male	7.4	21 (8.0 years)	Spelling 8

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

was "competitive" and didn't like getting things (work) wrong and often cried if she didn't "get the right answer." Given that this was an experiment where the proportion of error responses varied and that Alice found errors aversive, it was decided to withdraw Alice from the experiment and replace her with the next child from the pool of potential participants.

The experiments were conducted at the same three computer stations within the same school as Experiment 1. In Experiment 2 however, a screen was placed between each computer station to minimise interactions between participants. Fixed chairs also replaced swivel chairs to prevent participants swinging on the chairs. The new set up is shown in Figure 25.

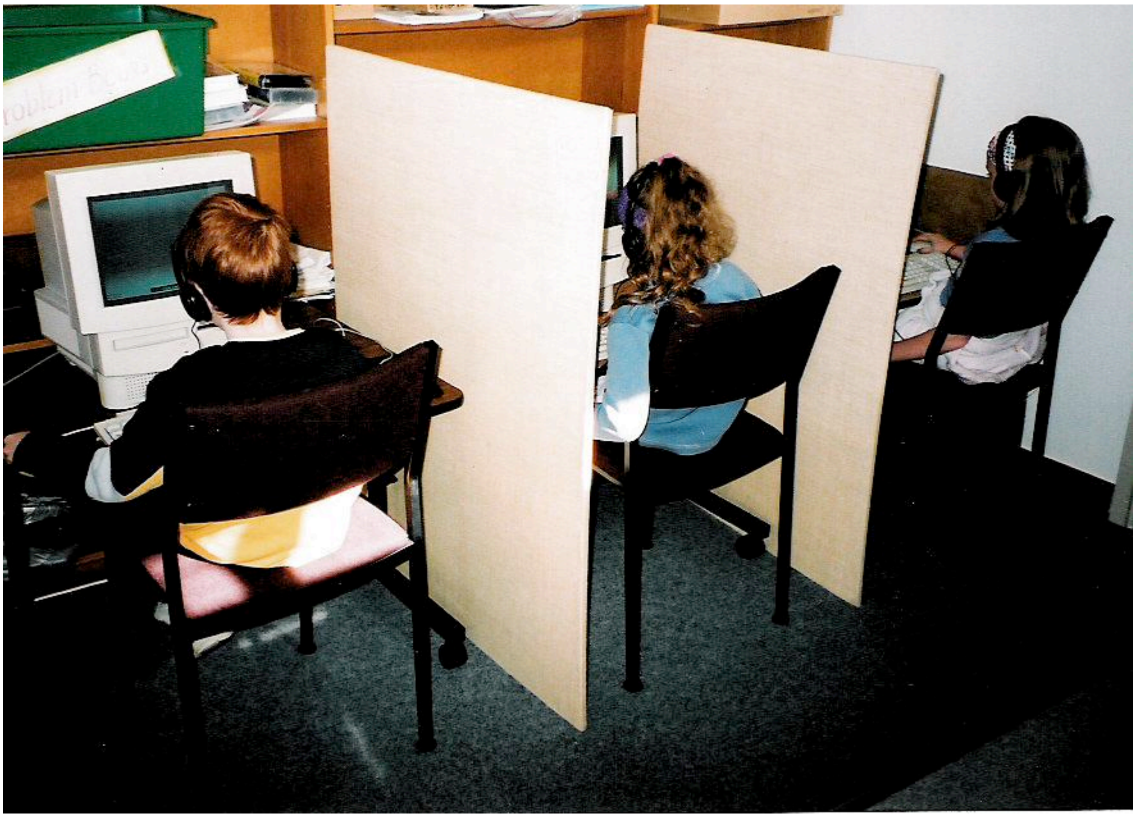


Figure 25. Experiment 2 participants working on the experimental spelling programme.

Learning Task

Spelling was selected as the learning task for the same reasons that it was selected in Experiment 1. As the participants were older than participants in Experiment 1 and their chronological reading ages were at least eight years old, it was decided to test them on the LYC Spelling 8 word list using the same procedure as testing for LYC Spelling 6 used in Experiment 1.

Pre-Experimental Procedures

The pre-experimental procedures were the same as Experiment 1. Children were tested on Spelling 8 words until 50 unknown words were located. They then provided with practice on the typing programme until they could type 20 correct letters per minute. Participants practised the spelling programme until they could operate it independently. Practice on the typing and spelling programme took two to three sessions. After practising the spelling programme the child was taught, and given the opportunity to play, some of the simple computer games which could be played for five minutes at the completion of experimental sessions.

Measurement Procedures

Twenty four hours after each practice session (and prior to the next day's practice session) the experimenter individually tested each participant for acquisition on either a 24-, 48- or 72-hour probe test on the first 10 words in a treatment. This was the same measurement procedure that was used in Experiment 1. In Experiment 2 however, only the original 10 words per treatment were measured for acquisition (words correct on 24-, 48- and 72-hour probe tests). Replacement words were however tested for 24-hour recall by asking the participant to spell the word. Words correct were removed from the practice set prior to the practice session and replaced with unknown words. This was done so that the accuracy level over 10 words could be calculated and controlled.

General Procedure

The general procedure for Experiment 2 was similar to Experiment 1. The difference was that the extraneous variable in Experiment 1 was removed. If the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at the end of the sentence as shown in Figure 26 and (b) the incorrect answer sound. No error-contingent model was presented.

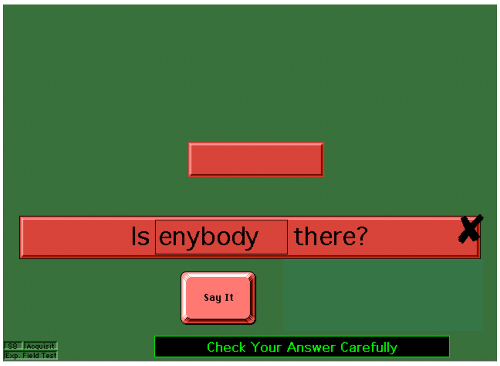


Figure 26. S8 completed response and error feedback screen in Experiment 2.

Experimental Design

A three-phase counter-balanced multiple-sequence across participants design was used. This is the same experimental design that was used in Experiment 1. However, in Experiment 2 each treatment concluded (a) once the participant had acquired the original 10 words, or (b) after 10 school days.

RESULTS

Interscorer Agreement

Accuracy checks were conducted by a second-year teacher trainee on the scoring of 25% of participants' responses on the probe tests. Interscorer agreement was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying this by 100. Interscorer agreement was 100%.

Procedural Reliability

A second-year teacher trainee conducted the procedural reliability check on 25% of the sessions to ensure that the participant received the correct level of modelling. Procedural reliability was assessed on agreements and disagreements between the within-session target accuracy level on the lesson printout and the experimenter's recording sheet of a session within each treatment. The procedural reliability was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying by 100. Procedural reliability was 100%.

Treatment Integrity

Treatment integrity was assessed by viewing the computer-lesson printouts showing the participants' responses and their within-session target accuracy levels on all words practised. The target accuracy level during instruction and the actual accuracy level during instruction of all words practised for each treatment for each participant can be seen in Table 12. Actual accuracy levels across all treatments ranged from 36% (Dean) to 97% (Carlos).

Table 12

Actual Accuracy Level During Instruction in Each Target Accuracy Level Treatment

	Actual accuracy level during instruction on all words		
	65% Target Accuracy Level Treatment	80% Target Accuracy Level Treatment	95% Target Accuracy Level Treatment
Adwin	65	69	84
Kate	69	79	88
Carlos	74	81	97
Patrick	70	86	95
Tania	72	86	78
Dean	59	66	36
Nicky	71	70	87
Casey	69	85	85
Bernard	69	65	77
Mean	69	76	81

The mean actual accuracy levels in the 65%, 80%, and 95% target accuracy level treatments were 69%, 76%, and 81%, respectively. For the 65% target accuracy level treatment, five participants achieved actual accuracy levels that were within ± 5% of the 65% target accuracy level, and the actual accuracy levels of all nine participants were within ± 10% of the target accuracy level. For the 80% target accuracy level treatment three participants achieved actual accuracy levels within ± 5% of the 80% target accuracy level, and the actual accuracy levels of six participants were ± 10%. For the 95% target accuracy level treatment two participants had actual accuracy levels which were within ± 5% of the 95% target accuracy level and five participants were within ± 10%. Actual accuracy levels were within ± 5% of the target accuracy level in 10 out of the 27 (37%) treatments, and were within ± 10% in 20 out of 27 (74%) treatments.

For four participants (Adwin, Kate, Carlos, Patrick), the lowest actual accuracy level was achieved in the 65% target accuracy level treatment, the next lowest actual accuracy level was achieved in the 80% target accuracy level treatment, and the highest actual accuracy level was achieved in the 95% target accuracy level treatment. That is, the 65% target accuracy level treatment produced the lowest actual accuracy level while the 95% target accuracy level treatment produced the highest actual accuracy level.

Treatment integrity was achieved for the 65% and 80% target accuracy level treatments as the mean actual accuracy levels were within ± 5%. However, treatment integrity was not achieved for the 95% target accuracy level treatment as the mean actual accuracy level was only 81%. So, while treatment integrity was not reached overall it was achieved to a higher degree than in Experiment 1.

Results

Rate of acquisition. Figure 27 shows the cumulative number of words correct on the 24-, 48- and 72-hour probe tests for each participant for each treatment. Words correct on the 72-hour probe test were classified as acquired. Twenty four, 48- and 72-hour probe data paths within treatments for each participant within each treatment were parallel in most cases. In almost all cases, words that were correct on the 24-hour probe test were also correct on the 72-hour probe test. Five of the participants acquired the first 10 words in all three treatments. Carlos and Tania acquired all 10 words in one treatment in one session. Carlos, Tania and Nicky acquired all 10 words in one treatment in two sessions. The total number of words acquired by all participants across all treatments was 262. The mean number of words acquired per participant was 29.1, and the mean number of words acquired per participant per treatment was 9.7.

Instructional efficiency. Mean trials to criterion for each participant within each treatment can be seen in Table 13. Across treatments for each participant the mean number of trials to criterion ranged from 3.0 for Tania and Carlos to 8.3 for Adwin. The mean number of trials to criterion was 5.9, 4.2 and 4.5 for the 65%, 80%, and 95% target accuracy level treatments, respectively. The mean number of trials to criterion across participants and treatments was 4.8 (SD = 1.5).

Effects of the model. The percentage of responses correct following the presentation of a model for the first round of a session (the first presentation of a word in a session) can be seen in Table 14. The effectiveness of the model in the first round ranged from 52% for Dean to 88% for Patrick. The mean percentage of correct responses following the model for the first round of a session was 78% compared to 61% in Experiment 1. It can be seen that the mean percentage of correct responses following the model was higher for acquired words (80%) than non-acquired words (65%). The percentage of correct responses following the model for all three rounds in a session ranged from 58% for Dean to 93% for Carlos and Patrick. The mean percentage of correct responses following the model for all three rounds of a session was 84% compared to 68% in Experiment 1.

Table 13 shows the actual accuracy level during instruction on the first 10 words practised (that is, the 10 words being measured for acquisition) within each treatment. This was calculated by adding the total number of correct responses of the first 10 words practised in a treatment and dividing this by the total number of practice responses on those words. This number was then converted it to a percentage. The mean accuracy levels on the first 10 words for each treatment was 71%, 79% and 83% for the 65%, 80% and 95% target accuracy level treatments, respectively. The variability between participants within treatments can be seen from Table 13. The mean actual accuracy level during instruction on the first 10 words practised across participants and treatments was 77%.

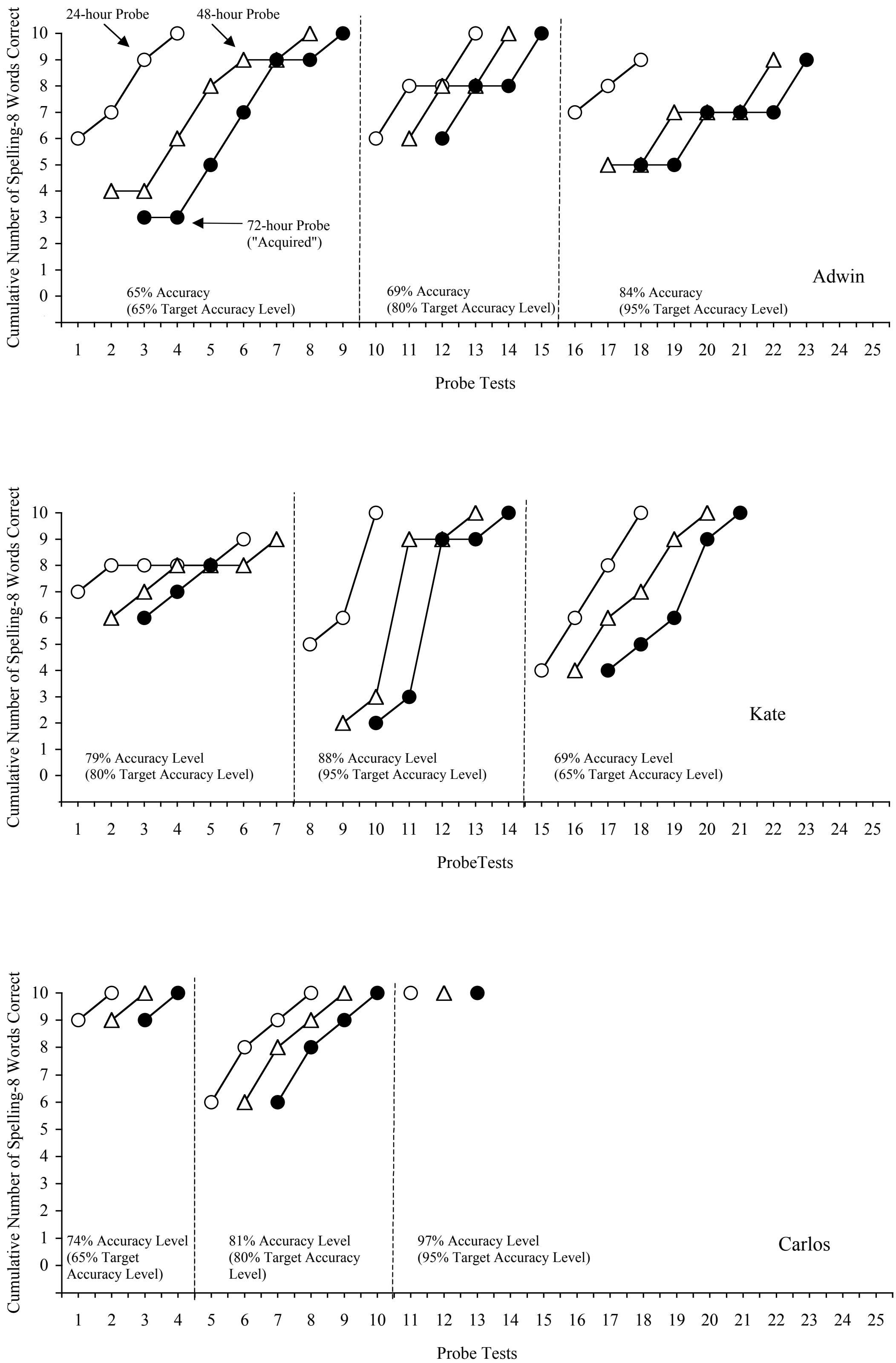
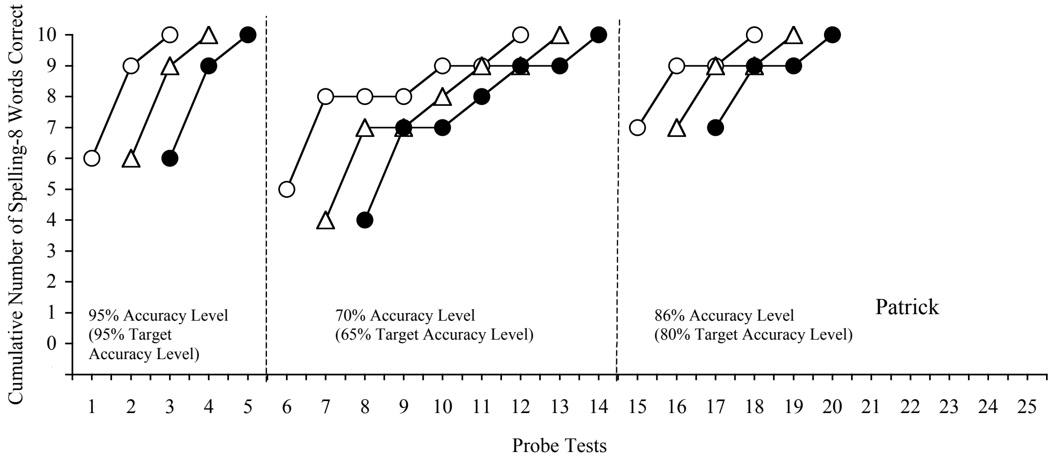
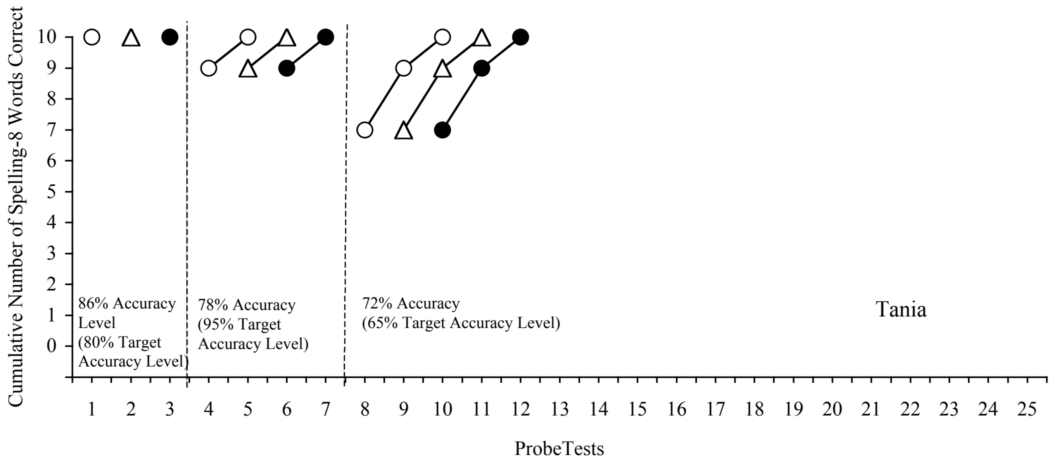


Figure 27. Cumulative number of words correct on the 24-, 48-, and 72-hour probe tests for each of the participants under each treatment for Experiment 2.





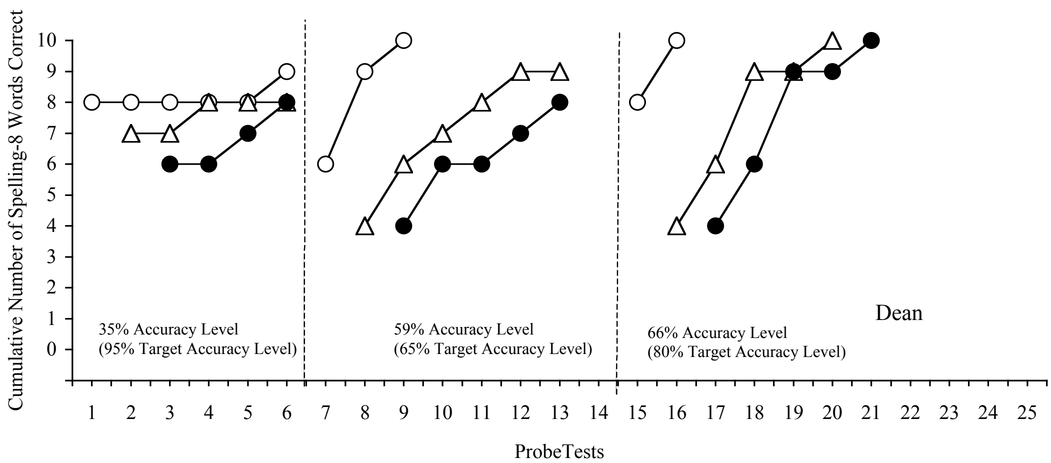


Figure 27. (Continued).

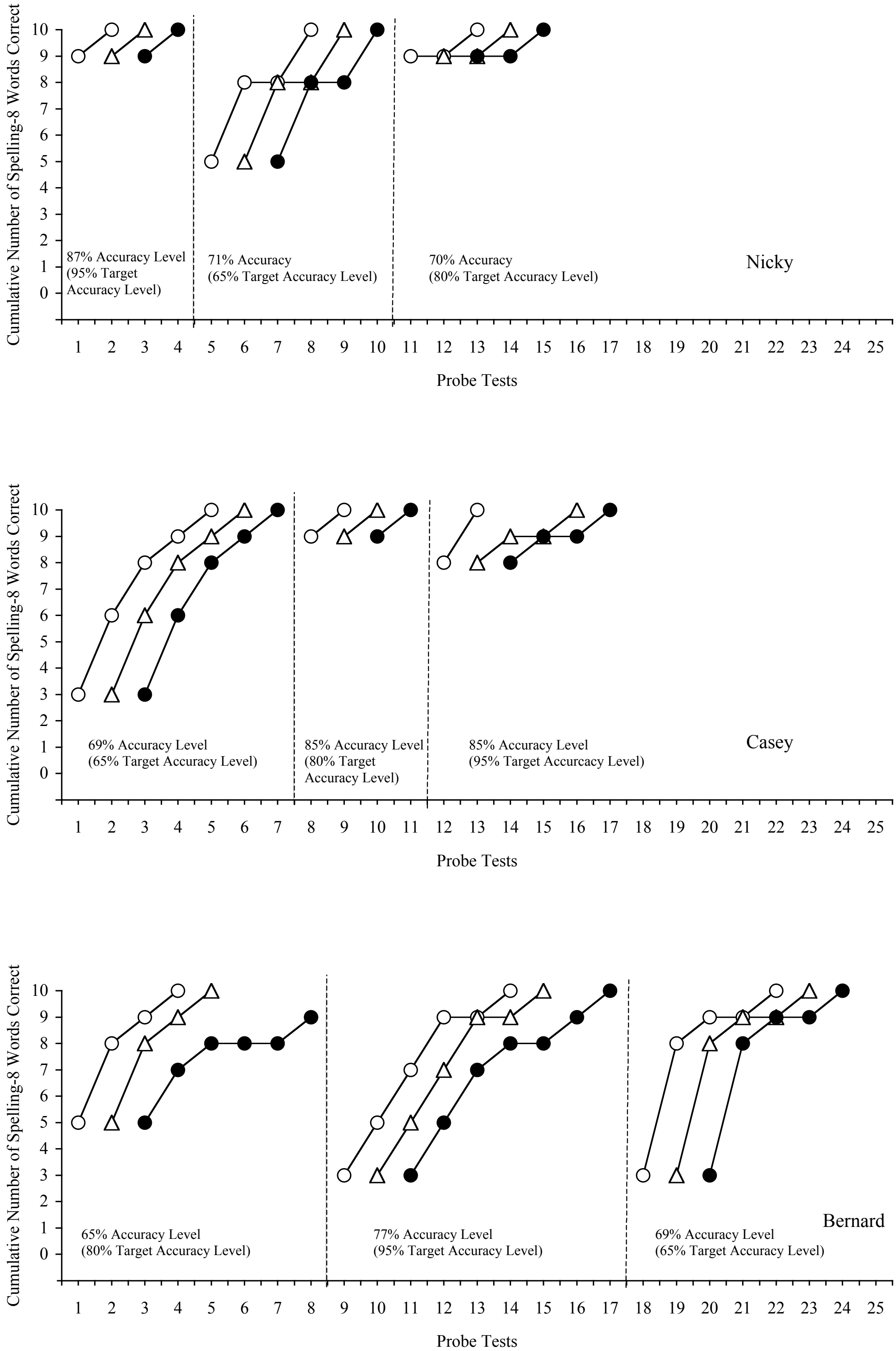


Figure 27. (Continued).

.

Table 13

Actual Accuracy Level During Instruction on the First 10 Words, Actual Accuracy Level on the Words Acquired, Mean Trials to Criterion for Each Participant Under Each Treatment, and Positive Relationships Between Actual Accuracy Level During Instruction and Trials to Criterion

	Treatment A			Treatment B			Treatment C			Positive relationship	
	65% Target Accuracy Level			80% Target Accuracy Level			95% Target Accuracy Level				
Participant	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words and mean trials to criterion	Actual accuracy level during instruction on words acquired and mean trials to criterion
Adwin	69	69	8.3	82	82	5.0	76	74	5.9	A-B Yes B-C Yes A-C Yes	A-B Yes B-C Yes A-C Yes
Kate	74	74	6.7	68	89	3.3	88	88	5.9	A-B No B-C No A-C Yes	A-B Yes B-C Yes A-C Yes
Carlos	88	88	3.3	76	76	5.0	97	97	3.0	A-B Yes B-C Yes A-C Yes	A-B Yes B-C Yes A-C Yes
Patrick	65	65	7.1	88	88	4.2	98	98	4.6	A-B Yes B-C No A-C Yes	A-B Yes B-C No A-C Yes
Tania	74	74	3.9	87	87	3.0	75	75	3.3	A-B Yes B-C Yes A-C Yes	A-B Yes B-C Yes A-C Yes
Dean	59	77	5.0	84	84	4.6	36[a]	77	3.7	A-B Yes B-C No A-C No	A-B Yes B-C No A-C No

Table 13

(Continued)

Participant	Treatment A 65% Target Accuracy Level			Treatment B 80% Target Accuracy Level			Treatment C 95% Target Accuracy Level			Positive relationship	
	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words	Actual accuracy level during instruction on words acquired	Mean trials to criterion	Actual accuracy level during instruction on first 10 words and mean trials to criterion	Actual accuracy level during instruction on words acquired and mean trials to criterion
Nicky	71	71	5.0	72	72	3.6	91	91	3.3	A-B Yes B-C Yes A-C Yes	A-B Yes B-C Yes A-C Yes
Casey	74	74	7.1	85	85	3.3	86	86	3.6	A-B Yes B-C No A-C Yes	A-B Yes B-C Yes A-C Yes
Bernard	64	64	6.3	71	69	5.3	52	52	7.1	A-B Yes B-C Yes A-C Yes	A-B Yes B-C Yes A-C Yes
Mean	71	73	5.9	79	81	4.2	83	82	4.5	-	-
SD	(8.3)	(7.1)	(1.7)	(7.6)	(7.3)	(0.9)	(15.0)	(14.4)	(1.5)	-	-
Total positive relationships	-	-	-	-	-	-	-	-	-	21	24

[a]Due to lack of treatment integrity, not included in mean total.

Table 14

Mean Percentage of Words Correct Following the Antecedent Model for First Presentation of a Word per Session and All Words per Session

Name	Mean percent of responses correct following model			
	During first presentation			During all presentations
	Acquired words	Non-acquired words	All words	All words
Adwin	83	100	85	87
Carlos	85	-	85	93
Tania	81	-	81	84
Nicky	81	-	81	87
Patrick	88	-	88	93
Casey	74	-	74	86
Dean	76	14	52	58
Kate	83	71	81	88
Bernard	69	75	71	78
Mean	80	65	78	84

Pair-wise comparisons were made to see whether a treatment that produced a higher actual accuracy level during instruction also produced the higher rate of acquisition when compared to a second treatment for each participant. Three comparisons were possible: comparisons between Treatment A and Treatment B, Treatment B and Treatment C, and Treatment A and Treatment C. The pair-wise analysis shows the relationship between accuracy level during instruction and rate of acquisition was positive in 21 (78%) of 27 cases. All three possible relationships for five of the participants were positive.

Table 13 also shows the actual accuracy level during instruction on the words acquired within each treatment. The actual accuracy level during instruction on the words acquired per treatment was calculated by adding the total number of correct responses on the words acquired in a treatment and dividing this by the total number of correct responses of the words acquired. The result was then converted to a percentage.

The mean accuracy level on the words acquired was 73%, 81% and 82% for the 65%, 80% and 95% target accuracy level treatments, respectively. The mean actual accuracy level during instruction on the words acquired across participants and treatments was 79%. The pair-wise analysis shows that out of 27 possible comparisons, 24 (89%) were positive. All three possible relationships for six of the participants were positive.

Social Validity

Social validity data was collected at the end of each treatment using the same procedure and questions as Experiment 1. As can be seen from Table 15, no participant reported in any of the treatments that they (a) did not like learning spelling on the computer or (b) found the spelling task "hard." Of the 27 treatments, 25 were reported as enjoyed "a lot" and two as "a bit." Task difficulty was reported as "easy" in 17 treatments and "middle" in 10.

DISCUSSION

In Experiment 2 participants acquired an average of 29.1 words each and required a mean of 4.8 trials to criterion on each of these words. This experiment demonstrated that the computer programme was effective for teaching spelling to young children.

Treatment Integrity

Two of the three treatments achieved mean actual accuracy levels within ±5% of the target accuracy level. This demonstrates that it is possible to control the accuracy level during instruction by removing error-contingent models, and it ensured that a high level of treatment integrity was achieved in Experiment 2. Eighty four percent of responses following the non-copying model were correct. There appear to be two main reasons why the antecedent model

Table 15

Participant's Social Validity Reports at the End of Experiment 2

	Enjoyment			Difficulty level		
	65% Target Accuracy Level	80% Target Accuracy Level	95% Target Accuracy Level	65% Target Accuracy Level	80% Target Accuracy Level	95% Target Accuracy Level
Bernard	A lot	A lot	A lot	Easy	Easy	Easy
Kate	A lot	A lot	A bit	Easy	Middle	Easy
Carlos	A lot	A lot	A lot	Easy	Middle	Easy
Patrick	A lot	A lot	A lot	Middle	Middle	Easy
Tania	A lot	A lot	A lot	Middle	Easy	Easy
Dean	A lot	A lot	A lot	Middle	Easy	Easy
Nicky	A lot	A lot	A lot	Middle	Middle	Easy
Casey	A lot	A lot	A lot	Middle	Easy	Easy
Adwin	A lot	A lot	A bit	Easy	Easy	Middle

exercised a high level of stimulus control over correct responding. First, the model occurred in only the antecedent position, and this seems to have resulted in the participants attending more carefully when the model of the correct spelling was presented. Second, the participants all had reading ages above their chronological age and appear to have learned to attend closely to such models. The one exception to this was Dean. During some sessions in Dean's first treatment (the 95% target accuracy level treatment) as soon as the stimulus question appeared on the screen Dean often pressed random letter keys and then the RETURN key.

Evaluation of the Experimental Procedures

Attending. The level of attention of the participants was better in Experiment 2 than in Experiment 1. Placing screens between the computer stations prevented participants from being distracted by each other. In the present experiment turning in the direction of another participant was not reinforcing because (a) it was difficult as the participant had to twist on a fixed chair and (b) it only produced the visual stimulus of the screen. When participants did talk to one another, they did not turn and face each other but kept looking at their own computer screens.

While the addition of screens and fixed chairs increased the overall attending behaviour of participants, when communication with other participants did occur, the intensity and obtrusiveness of the non-attending responses was high. Although occurring at low rates, participants still talked to one another. As there was a screen in place and they could not turn their body to talk to each other, they shouted instead. Occasionally, a participant would get out of his or her chair and walk to the next computer station to look at the other participant's screen. This sometimes occurred when a participant was playing a game and called another participant next to them and said, for example, "Look at this" or "Look where I am up to." When non-attending did occur, the experimenter redirected the participant to the spelling programme.

Task difficulty. The difficulty level of the words in each set was controlled by removing similar or known difficult-to-acquire words from the practice set prior to the experiment. However, as Spelling 8 words had only been used for one participant (Sade) in Experiment 1, many of the difficult-to-acquire words were not known prior to the experiment. This meant that there were still difficult-to-acquire words in the word sets for individual children. For example, the word *parents* was practised by seven participants. Two did not acquire it at all, and two others required more trials to criterion on this word than any other word. Uncontrolled variations in word difficulty still remained a problem in Experiment 2 because word sets still differed with respect to the number of total letter pairs which the participant needed to learn in order to acquire the set of words. For example, the pretest for Adwin showed that 23 letter pairs out of a total of 61 letter pairs were unknown in Set A and

22 letter pairs out of a total of 65 letter-pairs were unknown for Set B. However Set C had 40 unknown letter pairs out of a total of 67 letter pairs. Set C was therefore 40% more difficult for Adwin than the word sets for the other two treatments.

Relationship Between Accuracy Level During Instruction and Rate of Acquisition

The results in Table 13 show large numbers of positive relationships between the mean accuracy level during instruction and mean trials to criterion for each child. These positive relationships suggest that practice sessions were more efficient (that is, faster acquisition occurred) if participants were making correct responses rather than incorrect responses. These results are consistent with the results of the Chapter 4 literature review where it was hypothesised that this correlation occurs because participants have more opportunities to practise the correct response when instructional treatments are reasonably well prompted. However, as a classroom teacher, the experimenter is interested in investigating the effects of feedback variables on rate of acquisition following errors. This means that participants must generate some errors during instruction in subsequent experiments and there must be sufficient experimental control to ensure that these errors occur. A high level of experimental control over the accuracy level is therefore not possible because it will be recalled from Experiment 1 that presenting an antecedent model and some form of feedback variable (such as an error correction) created an experimental confound. In order to avoid this confound, a lower level of experiment control over the accuracy level during instruction will therefore be employed to ensure some errors occur in subsequent experiments. This can be achieved by manipulating the level of prompting provided by the antecedent models.

There were two main exceptions to the findings in Table 13 described above. For Dean the end-of-session game functioned as a reinforcer for progressing through the spelling programme as quickly as possible. This accounts for the very low accuracy level in his first experimental treatment (95% target accuracy level). Prior to the next treatment the experimenter put a contingency in place in which Dean had to attend to the spelling programme in order to gain access to the end-of-session game. During the 65% and 80% target accuracy level treatments (with the new contingency in place) Dean's results were consistent with those of the other participants.

A second exception was Patrick. The most probable reason for the lack of relationship between accuracy level during instruction and rate of acquisition in Patrick's case was poor control over task difficulty. During Treatment C, Patrick needed to acquire 33 letter pairs but during Treatment B he had only to acquire 26 letter pairs. So, while the accuracy level was higher during Treatment C (98%), Patrick needed to acquire 21% more letter pairs than he did during Treatment B where his accuracy level was 88%.

Conclusion

The present experiment found that it was possible to manage although not to completely control the accuracy level during instruction by a simultaneous, non-copying prompting procedure (Singleton, Schuster, Morse, & Collins, 1999). It was decided to lower the level of experimental control over the accuracy level in subsequent experiments when investigating feedback variables in order to avoid the experimental confound in Experiment 1. The experiment also found that words correct on the 24-hour probe test were almost always correct on the subsequent 48- and 72-hour probe test. Finally, Experiment 2 suggested that task difficulty would need to be better controlled in future experiments.

REFERENCES

Case, R. (1978). A developmentally based theory and technology of instruction. *Review of Educational Research, 48*, 439-463.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Massachusetts: Allyn and Bacon.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and application*. Eugene: ADI Press.

Nelly, E., & Smith, A. (2000). *PM Benchmark kit: An assessment resource for emergent - 12 years reading age*. Melbourne: Nelson ITP.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*, 51-67.

Singleton, D. K., Schuster, J. W., Morse, T. E., & Collins, B. C. (1999). A comparison of antecedent prompt and test and simultaneous prompting procedures in teaching grocery words to adolescents with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities, 34*, 182-199.

CHAPTER 6

RESEARCH INTO THE CORRECTION OF LEARNER ERRORS

The experiments in this thesis attempt to identify variables that affect rate of acquisition and instructional efficiency during the acquisition phase of learning for a discrete response such as spelling. As a classroom teacher the experimenter was particularly interested in the provision of feedback during classroom tasks.

Variables in the consequence position are important (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Grant & Evans, 1994) and one of the most important consequences for acquisition is feedback. Feedback is a stimulus presented to a learner, contingent upon a response, which provides information to the learner about one or more dimensions of the response (Grant & Evans, 1994).

Feedback effects rate of acquisition both during acquisition and during fluency building (Heward, 1994). In the classroom, a commonly used form of feedback is error correction. Error corrections take many forms as there are many different ways of responding to errors on learning tasks (Grimes, 1981; Heubusch & Lloyd, 1998; McCoy & Pany, 1986). Error corrections in spelling may be weak, as in the case of an error-contingent phonetic prompt which only prompts the initial sound of the word, or strong, as in the case of an error-contingent model of the correct response. Error corrections may model the correct response or they may refer to a rule. Elaborated error corrections can provide additional information. For example, if the correct response to “What is this?” is *cow*, elaborated error correction might also include additional information about cows. Error imitation is where the experimenter or the learner repeats the incorrect response in order to highlight the error component. Secondary response opportunities may follow an error correction. The learner is presented with an error-contingent model and then presented with the instructional stimulus again to provide a secondary-response opportunity. Another type of error-correction is one where the feedback accompanying the error correction is presented as a reprimand in the form of a “No” in a firm tone. Yet another type of error consequence is directed rehearsal. Directed rehearsal requires the learner to practise the correct response as part of the error-correction procedure. The learner is presented with an error-contingent model and is required to practise the correct response, say, five times while viewing the model. Error corrections may be presented immediately following the incorrect response, after a certain number of trials within the session, or may be presented at the end of the session. The teacher, the experimenter or a peer may present the error correction, or the learner may correct their own response by referring to, say, a spelling dictionary.

Several reviews of the effects of feedback and error-correction procedures have been undertaken. In a review of 20 experimental studies of feedback Walberg (1991) found a mean effect size 0.94 for corrective feedback. In a synthesis of 134 reviews measuring the effects of schooling Hattie (1992) found procedures such as reinforcement, remediation and feedback, and mastery learning produced effect sizes of 1.13, 0.65 and 0.50, respectively.

Getsie, Langer and Glass (1985) reviewed the effects of different combinations of feedback on children’s discrimination learning. Procedures that provided feedback on only correct responses produced an average effect size of –0.10, procedures that only corrected errors were found to have an average effect size of 0.20, and procedures that provided feedback on correct responses while at the same time correcting errors produced an average effect size of 0.24.

Two reviews (Heubusch & Lloyd, 1998; McCoy & Pany, 1986) examined the effects of different types of error correction during reading practice. The effects of feedback in the form of phonetic analysis, word supply, word meaning, sentence repeat, end-of-page review, and drill were examined. These procedures included within them differences in both the timing of feedback and the amount of additional practice required following an error.

Heubusch and Lloyd report that no single error correction procedure was superior while McCoy and Pany found word drill to be most effective.

Browder and Xin (1998), in a review of procedures used to teach and practice sight words, found that effective error-correction procedures included immediate error correction, and secondary-response opportunities.

The above reviews have shown that error-correction procedures can be effective in facilitating acquisition. Hattie (1992, p. 9) is of the view that “the most powerful single moderator that enhances achievement is feedback.”

Getsie et al. (1985) found that providing feedback for both correct and error responses was most effective and suggest that learners may be more motivated to avoid errors than to attain correct responses. Although not identified by Getsie et al. (1985), it seems that effective error-correction procedures share common characteristics. Walberg (1991)suggests that ideally teachers should “rapidly detect and remedy difficulties” by reteaching, or providing additional time or practice. Heubusch and Lloyd (1998) suggest that more successful error corrections are immediate and require the participant to engage in an active correct response following the correction. The superiority of the word drill procedure may be due to the additional practice responses that are generated (McCoy & Pany, 1986). These are the same findings as those of Browder and Xin (1998) who also found that procedures which were immediate and provided additional practices responses were more effective.

While the above reviews have shown that error correction facilitates acquisition, they have not identified those variables within the error-correction procedure that determine its effectiveness. As most error-correction procedures in the above reviews included additional practice opportunities, it is difficult to ascertain whether increased learning resulted from the type of error correction or whether it resulted from the additional error-contingent practice that often occurs following the error correction.

AIM

The present review had three aims:

1. To identify variables examined in the error-correction research.
2. To examine the effects of these variables on rate of acquisition.
3. To identify any shortcomings in the previous research which need to be avoided during further experimentation.

METHOD

Data based journals and empirical literature reviews listed in ERIC (1966 – 12/2002) and PsycINFO (1887 – 12/2002) were searched using the keywords "error correct*", "self correct*," "word supply," "word analysis," "consequen* prompt*," "contingent modeling," "modeling feedback," "prompt feedback," "correct* feedback," "performance feedback," "feedback and error*," "feedback and spelling," and "feedback and mathematic*." Search limits selected were "population = child*", "language = English", and "publication type = journal article." The tables of contents of the Journal of Applied Behavior Analysis and the Journal of Behavioral Education were searched for the last 10 years to identify additional reports that might meet the inclusionary criteria below.

Experiments were included in the present review if they met the following criteria:

1. The behaviours taught were discrete responses. Reading experiments were also included because error-correction procedures in reading attempt to correct a word (a discrete response) rather than the entire reading passage.

2. There were at least two different treatments comparing different procedures following learner errors.

3. The dependent variable was a measure of acquisition (not fluency or generalisation).

4. The learning task was an academic task such as spelling, mathematics, reading, writing, geography, or problem solving.

5. Participants were preschool or school-aged children.

Self-correction experiments were only included if the participant was provided with additional information following an error. Self-correction experiments where the learner made an error and self corrected without a prompt were not included.

Experiments in which the learner was provided with an additional response opportunity following an error were included. Directed rehearsal experiments (where more than one response opportunity after an error correction was provided) were only included if the overcorrection procedure was compared against an error-correction procedure. Reading studies where the participants read a new reading passage each session were excluded (because acquisition of new responses could not be shown).

The reference lists of each of the reports included in the review were also searched to identify further reports that might meet the criteria for inclusion.

Reports that met these criteria were coded in terms of participant characteristics, learning tasks, measures of learning, experimental design, independent variables, total number of response opportunities during practice, timing of feedback, and results.

Four experiments reported the overall results of the experiment and these are reported in the Mean Results column in Table 16. Twenty two experiments reported the mean results for each treatment. Where possible, these results were converted to a percentage and it is these percentages that are reported in the Mean Results column in Table 16. In the remaining 10 experiments either the percentage correct for each participant within each treatment or the number of trials to criterion for each participant within each treatment was examined. In almost all cases one treatment produced a higher rate of acquisition across all participants. It was decided to average the results of each treatment by (a) averaging the percentages across participants within a treatment or (b) by counting the number of trials to criterion for each participant for each treatment. These trials to criterion were aggregated across participants for each treatment and were converted to an average trials to criterion for that treatment. It is these average percentages per treatment or average trials to criterion per treatment that are reported in the Mean Results column in Table 16.

RESULTS

The search identified 32 studies (involving 36 experiments) which met the inclusionary criteria listed above. Table 16 describes the basic procedures and results of each of these experiments.

Research design. As can be seen from Table 16, five of the experiments employed between-group designs and 31 employed within-subject designs.

Participants and settings. Participants ranged from 6 to 17 years of age. Eight experiments classified their participants as normally developing, 26 as having some form of mental retardation or learning disability, and 2 as bilingual. Twenty-seven of the experiments took place in a preschool or school setting, one in an educational clinic, and one in a hospital. Seven studies did not describe the experimental setting.

Interobserver agreement. None of the five between-group experiments reported interobserver reliability data. Twenty-seven of the 31 single-subject experiments reported interobserver agreement and four did not.

Learning tasks. The learning tasks included learning spelling, reading, reading sight words, mathematical facts, learning geography facts, learning science vocabulary, learning reasoning skills, and sentence writing.

Independent variables. The 36 experiments examined the effects of a variety of error correction procedures. These included error-correction followed by a secondary-response opportunity, varying the strength of the error-contingent prompt, practising the correct

Table 16

Results of 36 Experiments Investigating the Effects of Various Error-Correction Procedures

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
Alvarado-Gomez and Belfiore (2000)	3 9-year old bilingual children	Learning spelling words	Number of words correct on 24 hour daily test	Alternating treatments	A: Practise word 3 times while viewing model B: Error-correction secondary-response	Number of response opportunities not reported	A: End of session B: Each response	A: 54% correct B: 71% correct
Barbetta and Heward (1993)	3 10- to 11-year olds with learning disabilities	Learning geography facts	Number of correct responses on immediate test (out of 7)	Alternating treatments	A: Error correction B: Error-correction secondary-response	A: 336 mean B: 496 mean	A: Each response B: Each response	A: 41% correct B: 63% correct
Barbetta, Heron, and Heward (1993)	6 8- to 9-year olds with developmental disabilities	Learning sight words	Number of words correct on immediate test (out of 10)	Alternating treatments	A: Error correction B: Error-correction secondary-response	A: 960 mean B: 1358 mean	A: Each response B: Each response	A: 50% correct B: 71% correct
Barbetta, Heward, and Bradley (1993)	5 8- to 9-year olds with developmental disabilities	Learning sight words	Number of words correct on immediate test (out of 7)	Alternating treatments	A: Whole-word error-correction secondary-response B: Phonetic-prompt error-correction secondary-response	A: 732 mean B: 831 mean	A: Each response B: Each response	A: 70% correct B: 47%correct
Barbetta, Heward, Bradley, and Miller (1994)	4 7- to 9-year olds with developmental disabilities	Learning sight words	Number of words correct on immediate test	Alternating treatments	A: Immediate error-correction secondary-response B: Immediate feedback & end-of-session error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response and end of session	Immediate error-correction secondary-response more effective

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
Bennett and Cavanaugh (1998) Experiment 1	9-year old classified as learning disabled	Learning multiplication facts	Number of responses correct on immediate daily test	Alternating treatments	A: No error correction B: Self error-correction secondary-response after every 4 trials (20 in total)	Number of response opportunities not reported	A: Not stated B: After every 4 trials	Higher accuracy on test with self-correction
Experiment 2					A: Self error-correction secondary-response after every 4 trials (20 in total) B: Self error-correction secondary-response after 20 trials	Number of response opportunities not reported	A: After every 4 trials B: After 20 trials	Higher accuracy on test with 4-trials self-correction
Carnine (1980)	9 normally developing 4- to 5-year olds	Learning sight words	Percent of words correct during training	Multiple baseline (no individual data)	A: Word supply error-correction secondary-response B: Phonetic prompt error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response	A: 20% correct B: 59% correct
Collins, Carnine, and Gersten (1987)	28 secondary school students classified as learning disabled	Learning reasoning skills on computer	Mean percent of responses correct on immediate posttest (out of 20)	Between groups	A: Error correction B: Elaborated error correction	Number of response opportunities not reported	A: Each response B: Each response	A: 57% correct B: 81% correct
Drevno et al. (1994)	5 9-year old children	Reading science words	Percent of correct definitions on 24-hour test	Alternating treatments	A: Error correction B: Error-correction secondary-response	A: 456 mean B: 761 mean	A: Each response B: Each response	A: 29% correct B: 41% correct
Espin and Deno (1989)	8 7- to 11-year old children classified as	Reading sight words	Percent of words correct at end of training	Multi element	A: Partial word prompt error-correction secondary-response	Number of response opportunities not	A: Each response	A: 68% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
	learning disabled				B: Whole word prompt error-correction secondary-response	reported	B: Each response	B: 100% correct
Gettinger (1993a)	65 normally developing 8-year old children	Learning spelling	Mean number of words correct on weekly test	Between groups	A: Words in sentences, study words on own B: Self error-correction secondary-response	Number of response opportunities not reported	A: Not stated B: End of session	A: 69% correct B: 87% correct
Gettinger (1993b)	4 normally developing 7-year olds	Learning spelling	Number of words correct on weekly test	Crossover design	A: End-of-session error correction B: Error-correction secondary-response	Number of response opportunities not reported	A: End of session B: Each trial	A: 45% correct B: 75% correct
Grskovic and Belfiore (1996)	5 10- to 11-year olds classified with emotional &/or learning disabilities	Learning spelling	Number of words correct on weekly test	Alternating treatments	A: Error-correction secondary-response B: Practise word 3 times while viewing model	Number of response opportunities not reported	A: Each trial B: End of session	A: 86% correct B: 70% correct
Hendrickson, Roberts, and Shores (1978)	2 primary school aged children with severe reading disabilities	Reading sight words	Number of words correct on immediate daily test	Multiple baseline multiple treatments	A: Antecedent prompt B: Error-correction secondary-response	Error correction > Antecedent by 283 error responses	A: Each response B: Each response	A: 70% correct B: 59% correct
Jenkins, Larson, and Fleisher (1983)	17 9- to 13-year olds classified as learning disabled	Reading	Mean percent of words correct on 24-hour daily test	Within subjects (no individual data)	A: Word supply error-correction secondary-response B: Word supply secondary-response error-correction & end-of-session drill	Number of response opportunities not reported, estimated drill > than other	A: Each response B: Each response	A: 47% correct B: 81% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant treatments	Timing of feedback	Mean results
Johnson, Schuster, and Bell (1996)	5 16- to 17-year olds classified as learning disabled	Learning science vocabulary	Trials to criterion	Alternating treatments	A: Simultaneous prompting B: Simultaneous prompting & error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response	A: 49 trials to criterion B: 66 trials to criterion
Kauffman, Hallahan, Haas, Brame, and Boren (1978) Experiment 1	2 8-year olds classified as mentally retarded	Learning spelling	Percent of responses correct on immediate test	Reversal design	A: Error-correction secondary-response B: Error imitation & error-correction secondary-response	Number of response opportunities not reported	A: Not stated B: Not stated	A: 29% correct B: 44% correct
Experiment 2	12-year old classified as learning disabled				A: Error-correction secondary-response B: Error imitation & error-correction secondary-response	Number of response opportunities not reported	A: Not stated B: Not stated	A: 83% correct B: 93% correct
McCurdy, Cundari, and Lentz (1990)	2 8- to 9-year olds classified as seriously emotionally disturbed	Reading sight words	Percent correct on 24-hour maintenance test	Multiple baseline	A: Progressive time delay & error correction B: Word supply error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response	A: 88% correct B: 85% correct
McGuffin, Martz, and Heron (1997)	6 8- to 9-year olds having difficulty with spelling	Learning spelling	Number of words correct on weekly test	Alternating treatments	A: Self error-correction secondary-response B: Practise word 5 times while viewing model	Number of response opportunities not reported, estimated practice 5 times	A: End of session B: Each response	A: 85% correct B: 51% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
						> self correct		
McNeish, Heron, and Okyere (1992)	5 13- to 14-year olds classified as learning disabled	Learning spelling	Number of words correct on weekly test	Alternating treatments	A: Writing words and studying words B: Self error-correction secondary-response	Number of response opportunities not reported	A: N/A B: End of list	A: 62% correct B: 86% correct
Meyer (1982)	58 7- to 14-year olds classified as learning disabled	Reading sight words	Mean number of responses incorrect on posttest (not stated when)	Between groups	A: Word supply error-correction secondary-response B: Phonetic prompt error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response	No significant differences
Morton, Heward, and Alber (1998)	5 11- to 12-year olds with learning disabilities	Learning spelling	Number of words correct on weekly test	Alternating treatments	A: Self error-correction secondary-response after every trial (10 in total) B: Self error-correction secondary-response after 10 trials	Number of response opportunities not reported	A: Each response B: After 10 responses	A: 63% correct B: 56% correct
Ollendick, Matson, Esveldt-Dawson, and Shapiro (1980) Experiment 2	2 12- to 13-year olds classified as learning disabled	Learning spelling	Number of 1^{st} responses correct during instruction (out of 8)	Alternating treatments	A: Directed rehearsal - practice error correctly 5 times B: Error-correction & 5 minutes study for all errors	Number of response opportunities not reported	A: End of session B: End of session	A: 58% correct B: 22% correct
Pany and McCoy (1988)	16 9-year olds classified as learning disabled	Reading	Mean number of words incorrect on immediate test	Repeated measures	A: No feedback B: Least-to-most prompt error correction	Number of response opportunities not reported	A: N/A B: Each response	A: 88% correct B: 92% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
Perkins (1988)	48 6- to 10-year olds classified as learning disabled	Reading	Mean number of words correct on immediate test	Between groups	A: No feedback B: Feedback & secondary response C: Word supply error-correction secondary-response D: Phonetic prompt error-correction secondary-response	Number of response opportunities not reported	A: N/A B: Each response C: Each response D: Each response	A: 38% correct B: 43% correct C: 99% correct D: 66% correct
Rosenberg (1986)	4 12- to 14-year olds classified as learning disabled	Reading	Percent of words correct on 24-hour later daily test	Alternating treatments	A: Word supply error-correction secondary-response B: Word supply error-correction secondary-response & end-of-session drill C: Word supply error-correction secondary-response & end-of-session phonic directed rehearsal	Number of response opportunities not reported, estimated error correction & drill > than error correction alone	A: Each response B: Each response C: Each response	A: 64% correct B: 86% correct C: 88% correct
Skinner, Shapiro, Turco, Cole, and Brown (1992)	6 normally developing 6-year olds	Learning multiplication problems	Number of problems correct during instruction (out of 12)	Alternating treatments	A: Self checking error-correction secondary-response B: Peer checking error-correction secondary-response	Number of response opportunities not reported	A: Each response B: Each response	A: 50% correct B: 50% correct
Stone and Serwatka (1982)	14-year old with learning disability	Learning to write sentences with correct	Percent of written words incorrect on 50-53 word passage	Reversal design (ABACA)	A: No feedback B: Oral error-correction secondary-response	A: Mean 50 R (26% errors) B: Mean 52 R (12% errors)	A: N/A B: End of session	A: 50% correct B: 88% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
		syntax			C: Oral self error-correction secondary-response	C: Mean 60 R (6% errors)	C: End of session	C: 94% correct
Van Houten (1993) Experiment 1	4 9- to 11-year olds classified as learning disabled	Subtraction problems	Percent of responses correct on 24-hour daily test	Alternating treatments	A: Whole answer error-correction secondary-response B: Rule error-correction secondary-response	Number of response opportunities not reported in any experiment	A: Each response B: Each response	A: 67% correct B: 93% correct
Experiment 2	2 6- to 8-year olds classified as learning disabled			Multiple baseline	A: Whole answer error-correction secondary-response B: Rule error-correction secondary-response		A: Each response B: Each response	A: 3% correct B: 98% correct
Experiment 3	2 6- to 7-year old boys classified as learning disabled			Multiple baseline	A: Whole answer error-correction secondary-response B: Rule error-correction secondary-response		A: Each response B: Each response	A: 26% correct B: 42% correct
Van Houten and Rolider (1989) Experiment 2	10 6- to 7-year old children having difficulty with number facts	Naming number facts	Percent of responses correct on 24-hour daily test	Alternating treatments	A: Error-correction secondary-response B: Error-correction secondary-response & reprimand for responding incorrectly	Number of response opportunities not reported	A: Each response B: Each response	A: 72% correct B: 77% correct
Vargas, Grskovic, Belfiore, and Halbert-Ayala (1997)	8 11- to 12-year old migrant students	Learning spelling	Number of words correct on 7-14 day posttest (total 24 words)	Alternating treatments	A: Practise word 3 times while viewing model B: Error-correction secondary-response	Number of response opportunities not reported	A: End of page B: Each response	A: 67% correct B: 92% correct

Table 16
(Continued).

Authors and Date	Participants	Task	Measure	Design	Independent variable	Number of response opportunities per participant	Timing of feedback	Mean results
Wirtz, Gardner, Weber, and Bullara (1996)	6 8- to 10-year olds with spelling difficulties	Learning spelling	Number of words correct on weekly test	Alternating treatments	A: Write the word 3 times and put into sentence B: Self error-correction secondary-response	Number of response opportunities not reported	A: N/A B: Each response	A: 53% correct B: 81% correct

response while viewing the model, latency of the error correction, self-correction, directed rehearsal, and reprimands followed by an error correction and a secondary-response opportunity.

DISCUSSION

The research examined in this review demonstrated that there are several error-correction procedures which are effective in increasing learning during instruction, and that there are a variety of variables within these error-correction procedures which may function to affect rate of acquisition.

Five experiments (Bennett & Cavanaugh, 1998, Experiment 1; Johnson et al., 1996; Pany & McCoy, 1988; Perkins, 1988; Stone & Serwatka, 1982) compared the effects of an error-correction procedure against no error correction. One of these experiments (Johnson et al., 1996) compared an error correction and a secondary response procedure against no error corrections under well prompted conditions. All five experiments found the error-correction procedure to be more effective than no error correction. These results seem logical given that an error is likely to re-occur if the learner receives no additional information regarding how to respond correctly. As all of the experiments required the participants to produce a secondary response it is impossible to determine whether the difference in rate of acquisition was a function of the error-correction procedure (that is, the presentation of the correct response) or whether it was a function of the additional practice that participants received, or both.

Two experiments (Hendrickson et al., 1978; McCurdy et al., 1990) compared error correction with an antecedent procedure. Results from both favoured the antecedent-prompting procedure. It is likely the higher accuracy level during instruction generated by the antecedent-prompt procedure affected rate of acquisition.

Three experiments (Barbetta, Heron et al., 1993; Barbetta & Heward, 1993; Drevno et al., 1994) found that providing a secondary-response opportunity after an error correction was more effective than providing an error correction alone. However, in all secondary response-opportunity treatments the secondary response was an additional response within each trial. It seems probable that a treatment that provides additional practice responses will be more effective.

Seven experiments (Alvarado-Gomez & Belfiore, 2000; Gettinger, 1993a; Grskovic & Belfiore, 1996; McGuffin et al., 1997; McNeish et al., 1992; Vargas et al., 1997; Wirtz et al., 1996) compared error correction against a traditional teaching method in spelling. All seven experiments found word-by-word error correction to be more effective. The traditional method requires the participant to copy a model a number of times and may involve studying the word, putting it into sentences or looking it up in the dictionary. The unstructured nature of the traditional teaching method means that different participants will have received different numbers of response opportunities. This means none of these experiments controlled the number of response opportunities. It is probable that error-correction treatments focus the learner's attention on the unknown letter sequence whereas the traditional method merely requires a copying response – a condition known to yield little learning (Kulhavy, Yekovich, & Dyer, 1976).

Four experiments compared the effects of an immediate error correction against a delayed error correction (Barbetta et al., 1994; Bennett & Cavanaugh, 1998, Experiment 2; Gettinger, 1993b; Morton et al., 1998). All four found the immediate error correction more effective. However, in all treatments the learner was required to respond again after the error correction. The error-contingent model therefore functioned as a prompt for a secondary response that generated additional correct responses during practice.

Two experiments (Skinner et al., 1992; Stone & Serwatka, 1982) compared the effects of the participant providing their own error correction with the experimenter providing the error correction. Both experiments found that self error-correction was overall more effective. However, the error-correction procedures contained an extraneous variable in Stone and Serwatka (1982). The participant, who could say a sentence with correct syntax, was learning to write a sentence with correct syntax. In one treatment, after the participant had written a sentence, she was asked, "What is wrong with this sentence?" If the participant could not self-correct the experimenter explained the reason for the error. The participant then corrected the written response. In the other treatment, all incorrect sentences were read into a tape recorder and played back to the participant one at a time. The participant then played back the tape and read her written responses. This allowed the participant to hear the incorrect response in the form of an oral response. The participant corrected an error after it was heard on the tape by rewriting the error component of the sentence. It appears this procedure functioned as a strong prompt to correct any incorrectly written sentences. However, it cannot be determined if the superior effects of this treatment were due to the experimenter or the participant correcting the error, or whether they were due to the explanation of the error or presentation of the error in an oral form. Self-correction was more effective than peer correction for four out of six participants in Skinner et al. (1992), although overall mean results were the same.

Twelve experiments (Barbetta, Heward et al., 1993; Carnine, 1980; Collins et al., 1987; Espin & Deno, 1989; Kauffman et al., 1978, Experiment 1 & 2; Meyer, 1982; Perkins, 1988; Van Houten, 1993, Experiments 1, 2 & 3; Van Houten & Rolider, 1989, Experiment 2) compared the effects of varying the strength of the error-contingent prompt. None of these experiments controlled for response opportunities and ten found the stronger prompt more effective. One experiment (Carnine, 1980) found the weaker prompt more effective. This was a reading study comparing word supply with some form of phonic analysis. It is possible the weaker prompts were more effective because they taught the learner a decoding strategy that could be applied to other unknown words. Another reading experiment (Meyer, 1982) found no significant difference between a word analysis and word supply treatments. Meyer suggests that the word-analysis procedure may have been too complicated for the participants as they were classified as learning disabled, and this reduced the effectiveness of the procedure.

Three experiments (Jenkins et al., 1983; Ollendick et al., 1980; Rosenberg, 1986) measured the effects of directed rehearsal following an error correction. Directed rehearsal requires the learner to emit the correct response, say, five times. All three found the directed-rehearsal treatment to be more effective than an error correction or error-correction secondary-response treatment. While the response opportunities were not reported or controlled in any of these experiments it could be inferred that the directed-rehearsal treatments produced considerably higher correct responses during practice. It is possible these additional correct practice responses affected rate of acquisition. Another possibility is that the directed-rehearsal procedure functioned as an aversive consequence which motivated the learners to take more care to avoid errors.

Weaknesses of the Experiments

Response opportunities. A major weakness of all 36 experiments in this review is that the number of response opportunities was not controlled. In 31experiments the numbers of response opportunities were not even reported. If the number of response opportunities is not controlled it is impossible to tell whether a difference in treatment effects is the result of the independent variable or whether it is due to variations in secondary-response opportunities.

Interobserver reliability. A weakness of all of the between-groups and four of the within-subject studies was the absence of data on interobserver agreement.

Conclusion

The present review identified variables examined in the error-correction research. These included practising the correct response while viewing the model, varying the strength of error-contingent prompt, latency of the error correction, self-correction, error correction followed by a secondary-response opportunity, reprimands followed by an error correction and a secondary-response opportunity, and directed rehearsal.

This review has shown that error-correction procedures result in greater acquisition than procedures such as no feedback, feedback alone and, in the case of spelling, the traditional teaching method. However, when compared to an antecedent-prompt treatment, error correction was reported to be less effective.

The most effective error-correction procedures appear to be those that include self-correction, immediate error correction, providing a strong prompt with the error correction, and providing a secondary-response opportunity.

However, none of the 36 experiments controlled the number of response opportunities across treatments. In other words, the treatment comparisons are difficult to interpret because it is not known whether the observed effects are a function of the independent variable or a function of the number of response opportunities occurring during practice.

Given that all of the 36 experiments reviewed had uncontrolled extraneous variables operating, it may be unwise to draw conclusions from these experiments. While this review has found that error corrections can result in faster acquisitions, it has failed to identify the variables responsible for this effect.

REFERENCES

Alvarado-Gomez, M., & Belfiore, P. J. (2000). Effects of error correction on Spanish spelling words for students whose primary language is Spanish. *Journal of Behavioral Education, 10*, 213-221.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition and maintenance of geography facts by elementary students with learning disabilities. *Journal of Behavioral Education, 3*, 217-233.

Barbetta, P. M., Heward, W. L., & Bradley, D. M. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110.

Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A. D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 27*, 177-178.

Bennett, K., & Cavanaugh, R. A. (1998). Effects of immediate self-correction, delayed self-correction, and no correction on the acquisition and maintenance of multiplication facts by a fourth-grade student with learning disabilities. *Journal of Applied Behavior Analysis, 31*, 303-306.

Browder, D. M., & Xin, Y. P. (1998). A meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education, 32*, 130-153.

Carnine, D. W. (1980). Phonic versus whole-word correction procedures following phonic instruction. *Education and Treatment of Children, 3*, 323-330.

Collins, M., Carnine, D. W., & Gersten, R. (1987). Elaborated corrective feedback and the acquisition of reasoning skills: A study of computer-assisted instruction. *Exceptional Children, 54*, 254-262.

Drevno, G. E., Kimball, J. W., Possi, M. K., Heward, W. L., Gardner, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and generalization of science vocabulary by elementary students: A systematic replication. *Journal of Applied Behavior Analysis, 27*, 179-180.

Espin, C. A., & Deno, S. L. (1989). The effects of modeling and prompting feedback strategies on sight word reading of students labeled learning disabled. *Education and Treatment of Children, 12*, 219-231.

Getsie, R. L., Langer, P., & Glass, G. V. (1985). Meta-analysis of the effects of type and combination of feedback on children's discrimination learning. *Review of Educational Research, 55*, 9-22.

Gettinger, M. (1993a). Effects of error correction on third graders' spelling. *Journal of Educational Research, 87*, 39-45.

Gettinger, M. (1993b). Effects of invented spelling and direct instruction on spelling performance of second-grade boys. *Journal of Applied Behavior Analysis, 26*, 281-291.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: HarperCollins.

Grimes, L. (1981). Error analysis and error correction procedures. *Teaching Exceptional Children, 14*, 17-21.

Grskovic, J. A., & Belfiore, P. J. (1996). Improving the spelling performance of students with disabilities. *Journal of Behavioral Education, 6*, 343-354.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*, 5-13.

Hendrickson, J., Roberts, M., & Shores, R. E. (1978). Antecedent and contingent modeling to teach basic sight vocabulary to learning disabled children. *Journal of Learning Disabilities, 11*, 69-73.

Heubusch, J. D., & Lloyd, J. W. (1998). Corrective feedback in oral reading. *Journal of Behavioral Education*, 63-79.

Heward, W. L. (1994). Three "low tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283-320). California: Brooks/Cole Publishing Company.

Jenkins, J. R., Larson, K., & Fleisher, L. (1983). Effects of error correction on word recognition and reading comprehension. *Learning Disability Quarterly, 6*, 139-145.

Johnson, P., Schuster, J., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6*, 437-458.

Kauffman, J. M., Hallahan, D. P., Haas, K., Brame, T., & Boren, R. (1978). Imitating children's errors to improve their spelling performance. *Journal of Learning Disabilities, 11*, 33-38.

Kulhavy, R., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522-528.

McCoy, K. M., & Pany, D. (1986). Summary and analysis of oral reading corrective feedback research. *Reading Teacher, 39*, 548-554.

McCurdy, B. L., Cundari, L., & Lentz, F. E. (1990). Enhancing instructional efficiency: An examination of time delay and the opportunity to observe instruction. *Education and Treatment of Children, 13*, 226-238.

McGuffin, M. E., Martz, S. A., & Heron, T. E. (1997). The effects of self-correction versus traditional spelling on the spelling performance and maintenance of third grade students. *Journal of Behavioral Education, 7*, 463-476.

McNeish, J., Heron, T. E., & Okyere, B. (1992). Effects of self-correction on the spelling performance of junior high students with learning disabilities. *Journal of Behavioral Education, 2*, 17-27.

Meyer, L. A. (1982). The relative effects of word-analysis and word-supply correction procedures with poor readers during word-attack training. *Reading Research Quarterly, 17*, 544-555.

Morton, W. L., Heward, W. L., & Alber, S. R. (1998). When to self-correct? A comparison of two procedures on spelling performance. *Journal of Behavioral Education, 8*, 321-335.

Ollendick, T. H., Matson, J. L., Esveldt-Dawson, K., & Shapiro, E. S. (1980). Increasing spelling achievement: An analysis of treatment procedures utilizing an alternating treatments design. *Journal of Applied Behavior Analysis, 13*, 645-654.

Pany, D., & McCoy, K. M. (1988). Effects of corrective feedback on word accuracy and reading comprehension of readers with learning disabilities. *Journal of Learning Disabilities, 21*, 546-550.

Perkins, V. L. (1988). Feedback effects on oral reading errors of children with learning disabilities. *Journal of Learning Disabilities, 21*, 244-248.

Rosenberg, M. S. (1986). Error-correction during oral reading: A comparison of three techniques. *Learning Disability Quarterly, 9*, 182- 192.

Skinner, C. H., Shapiro, E. S., Turco, T. L., Cole, C. L., & Brown, D. K. (1992). A comparison of self- and peer-delivered immediate corrective feedback on multiplication performance. *The Journal of School Psychology, 30*, 101-116.

Stone, A. K., & Serwatka, T. S. (1982). Reducing syntactic errors in written responses of a retarded adolescent through oral patterning. *Education and Training of the Mentally Retarded, 17*, 71-74.

Van Houten, R. (1993). Rote vs. rules: A comparison of two teaching and correction strategies for teaching basic subtraction facts. *Education and Treatment of Children, 16*, 147-159.

Van Houten, R., & Rolider, A. (1989). An analysis of several variables influencing the efficacy of flash card instruction. *Journal of Applied Behavior Analysis, 22*, 111-118.

Vargas, A. U., Grskovic, J. A., Belfiore, P. J., & Halbert-Ayala, J. (1997). Improving migrant students' spelling of English and Spanish words with error correction. *Journal of Behavioral Education, 7*, 13-23.

Walberg, H. J. (1991). Productive teaching and instruction: Assessing the knowledge base. In H. C. Waxman & H. J. Walberg (Eds.), *Effective teaching: Current research*. Berkeley: McCutchan Publishing Corporation.

Wirtz, C. L., Gardner, R., Weber, K., & Bullara, D. (1996). Using self-correction to improve the spelling performance of low-achieving third graders. *Remedial and Special Education, 17*, 48-58.

CHAPTER 7

THE EFFECTS OF MODELS AS PROMPTS AND AS ERROR CORRECTIONS

In the classroom, teacher guidance that the learner receives can vary from extensive and well structured to almost nothing. Asking a learner to complete a task with no support is often referred to as trial-and-error learning (Ellis, Ludlow, & Walls, 1978). It is possible to learn under trial-and-error conditions provided a learner's response either (a) generates it own feedback or else (b) is followed by a stimulus that informs the learner of the correctness of the response, and provides some form of correction if the response is incorrect (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Grant & Evans, 1994).

The provision of a prompt in the form of a model has been found to accelerate acquisition of a variety of academic skills such as learning spelling (e.g., Stevens & Schuster, 1987), learning mathematics (e.g., Koscinski & Gast, 1993), learning science facts (e.g., Johnson, Schuster, & Bell, 1996), naming letters of the alphabet (e.g., Griffiths & Griffiths, 1976), reading sight words (e.g., Browder & Lalli, 1991; Espin & Deno, 1989; Gast, Ault, Wolery, & Doyle, 1988), and reading (e.g., McGee & McCoy, 1981). In Experiment 1, however, the 6-year old learners responded in different ways to the antecedent model (the spelling of the word). Some learners vocalized the letters of the word, some attended briefly to the model, and others appeared to not attend to the model before responding.

The presentation of a model of the correct spelling of a word as a consequence of an incorrect response has also been shown to accelerate the acquisition of academic skills such as learning spelling (e.g., Simonsen & Gunter, 2001), learning mathematics (e.g., Bennett & Cavanaugh, 1998), learning science words (e.g., Drevno et al., 1994), learning letters of the alphabet (e.g., Trap, Milner-Davis, Joseph, & Cooper, 1978), reading sight words (e.g., Barbetta, Heron, & Heward, 1993) and reading (e.g., Heubusch & Lloyd, 1998). In Experiment 1 individual learners also responded differently to the presentation of a model (the correct spelling of the word) following an incorrect response. Some did not attend to the model and immediately moved to the next trial. Others attended to the feedback before moving to the next trial, and some rehearsed only those parts of their response that had been incorrect before moving to the next trial.

Several reviews (Aiken & Lau, 1967; Ault, Wolery, Doyle, & Gast, 1989; Browder & Xin, 1998; Lysakowski & Walberg, 1982) have compared the relative effects of some form of antecedent model versus some form of consequent model. An early review by Aiken and Lau (1967) compared the effects of antecedent modelling versus error correction in paired-associate, perceptual, and discrimination tasks. They concluded that in "many instances, the two procedures result in about the same level of performance" (Aiken & Lau, 1967, p. 339). Tasks were typically non-classroom type activities (e.g., identifying code lines) and it is likely that most of the participants in the experiments were adults.

Lysakowski and Walberg (1982) undertook a meta-analysis of experiments examining the effectiveness of some form of prompting and studies investigating some form of error correction. In a review of 54 experiments they found that both procedures produced "robust, consistent, and moderately large effects" and that the differences in the size of the effect produced by the two procedures were not significant.

Ault et al. (1989) reviewed four experimental evaluations of prompting procedures and error-correction procedures. The learning task in all studies was a chained motor task. Ault et al. (1989, p. 352) concluded that antecedent-prompting procedures "that have been compared to trial-and-error and error-correction procedures are more effective and . . . more efficient." Given the superiority of antecedent-modelling procedures Ault et al. concluded that there appears to be "little justification for using trial-and-error or error-correction procedures." These results however are based on studies where the participants were classified as either moderately or severely handicapped and the authors acknowledge that more replications are needed across learners with differing levels of ability. In addition, the authors state that prompting procedures require more planning and are procedurally more complex to administer, especially in the classroom.

In a meta-analysis of procedures for teaching sight words Browder and Xin (1998) found that error-correction procedures produced “better outcomes” than antecedent-prompting procedures. The authors concluded that error-correction procedures may be more effective than previously assumed. They also argued that postresponse procedures have become more sophisticated and often include other variables in addition to the prompt such as requiring additional responding from the learner in the form of rapid drill, response repetition, and active student responding.

To summarise, literature reviews to date have concluded that error correction is as effective as prompting procedures during the acquisition phase. Only one (Browder & Xin, 1998) found that error-correction procedures were more effective than antecedent-prompting procedures. However, it still remains unclear which procedure is more effective. A review was therefore undertaken in an attempt to answer this question.

Literature Review

Data based journals and empirical literature reviews listed in PsycINFO (1887 – 10/2003) were searched using the keywords “errorless or preresponse or prompt* and feedback or postresponse or trial and error or error correct* and PO = childhood. Research articles from the accuracy level literature review in Chapter 5 were also included if they met the criteria below.

Studies were included in the present review if (a) there were at least two different treatments, (b) one treatment employed an antecedent prompt procedure and the other treatment employed a consequent prompt procedure, (c) the behaviours taught were discrete responses, and (d) the behaviours were academic, concept learning or discrimination type tasks (motor skill or chained tasks were not included).

The tables of contents of the Journal of Applied Behavior Analysis, the Journal of Behavioral Education, the Journal of Learning Disabilities and Research in Developmental Disabilities were searched for the last ten years to identify reports that might meet the inclusionary criteria below. The reference lists of each of the research articles included in the review were also searched to identify further reports that might meet the criteria for inclusion.

Ten experiments of the effects of the relative contribution of antecedent-modelling versus error-correction procedures on rate of acquisition with children were located. These are summarized in Table 17.

Research design. The participants in these 10 experiments ranged from preschoolers to 10-year olds. Six experiments classified the participants as normally developing, and two as having some form of disability. Two experiments compared normally-developing participants and participants with developmental delays. All experiments took place in school settings. Three of the seven within-subjects experiments (Hendrickson, Roberts, & Shores, 1978; McCurdy, Cundari, & Lentz, 1990; Smeets, 1992) provided interobserver reliability data. None of the three between-groups experiments reported interobserver reliability data for either the responses made during the experiment or the responses made on the posttest. Response opportunities were controlled in only one between-groups study (Egeland, 1975). The remaining experiments either did not control or did not report the number of response opportunities during practice.

Effect sizes. Where possible, effect sizes for within-subject and between-group experiments were calculated using Glass, McGaw and Smith's (1981) formula:

$$\frac{\text{(mean score of Treatment A – mean score of Treatment B)}}{\text{standard deviation of the Treatment B}}$$

The Antecedent-Model Treatment was assigned as Treatment A and the Error-Correction Treatment was assigned as Treatment B for both between-group and within-subject experiments. For within-subject experiments, where only the mean participant score was presented for each treatment (that is, no session-by-session data was presented), the mean scores of all participants in a treatment were added and divided by the number of participants. This produced a mean score per treatment. For within-subject experiments, where session-by-session data was presented, only the scores of the last three sessions for each participant were inspected as suggested by Swanson and Sachse-Lee (2000). The final three session scores for all participants in a treatment were added and divided by the total number of participants for that treatment. This produced a mean score per treatment. The effect size was calculated on the trials to criterion in Griffiths and Griffiths (1976). In this experiment the lower the score the more effective the treatment. So that a negative effect size was not produced on the most effect treatment the negative effect size was reported as a positive effect size for this experiment.

Results. Table 17 shows that 9 of the 10 experiments reported that the antecedent-model treatment was more effective than the error-correction treatment. One study (Ellis et al., 1978) reported that the error-correction procedure was more effective. All experiments in which an effect size could be calculated favoured the antecedent-model treatment. The effect size could be calculated on posttest performance, number or percent responses correct, or trials to criterion on five experiments. The mean effect size favouring the antecedent model across the five experiments was 1.9 (SD = 1.4).

Table 17

Results of 10 Experiments Comparing the Effects of Antecedent Modelling and Error-Contingent Modelling Procedures

Authors and date	Participants	Task	Measure	Design	Independent variable	Number of total response opportunities per participant for each treatment	Effect size	Results
Duffy and Wishart (1987)	8 7- to 9-year olds with Down's Syndrome	Shape discrimination	Number of responses correct	Within subjects	A: Fading B: Error correction	Response opportunities not controlled	1.9	Overall, errorless more effective both in training and posttest for both groups
	8 normally developing preschoolers						1.3	
Egeland (1975)	108 Preschoolers	Naming letters of alphabet	Number of letters correct	Between groups	A: Errorless B: Error correction	A: 60 responses B: 60 responses	-[a]	Errorless procedures more effective
Ellis, Ludlow, and Walls (1978)	27 4th grade students	Naming Kanji symbols	Percent errors	Between groups	A: Fading B: Error correction	More response opportunities for fading	-	Fading took more trials to criterion
Everett (1977)	108 2nd grade children	Identifying line tilts and dots	Mean number of errors	Between groups	A: Errorless B: Error correction	No of response opportunities not reported	-	Errorless more effective
Griffiths and Griffiths (1976)	6 normally developing 5- to 6-year old children	Discriminating between "b,d" & "p,q"	1: Trials to criterion 2: Total errors	Within subjects counter balanced across tasks	A: Stimulus fading B: Error correction	Response opportunities not controlled	4.3	Fewer errors and trials to criterion with fading
Gollin and Savoy (1968)	52 normally developing 3- to 8-year old children	Discriminating shapes	Number of responses correct	Within subjects	A: Fading B: Error correction	Response opportunities not controlled	-	Fewer errors with fading during training
Hendrickson, Roberts, and Shores (1978)	2 primary school aged children with severe	Reading sight words	Number of words correct on immediate daily	Multiple baseline multiple treatments	A: Antecedent model B: Error-correction	Error correction > antecedent by 141 mean error	1.8	A: 18 correct (mean) B: 15 correct (mean)

Authors and date	Participants	Task	Measure	Design	Independent variable	Number of total response opportunities per participant for each treatment	Effect size	Results
	reading disabilities		test		secondary-response	responses		
McCurdy, Cundari, and Lentz (1990)	2 8- to 9-year olds classified as seriously emotionally disturbed	Reading sight words	1. Number of responses correct during instruction 2. Percent correct on 1-day maintenance test	Multiple baseline	A: Time-delay secondary-response B: Error-correction secondary-response	No of response opportunities not reported	0.4	A: 74% correct during instruction (mean) -1 day test 88% correct (mean) B: 66% correct during instruction (mean) -1 day test 85% correct (mean)
Robinson and Storm (1978)	18 normal 6-year olds	Colour discrimination	Number of errors during acquisition	Within-subject	A: System-of-most prompts B: Error correction	Response opportunities not controlled	-	Less errors for system-of-least prompts and better maintenance
Smeets (1992)	30 normally developing 4- to 5-year olds 30 6- to 10-year olds with mildly mentally retarded 30 4- to 6-year olds with moderate mental retardation	Mirror image discrimination	Trials to criterion	Between groups	A: Time delay B: Error correction	Each treatment was stopped if criterion not reached after 40 practice responses	-	Time delay more effective for all populations

[a] Calculating effect size was not possible..

Discussion. The results of Experiment 2 (Chapter 5) showed that there was a relationship between the accuracy level during instruction and rate of acquisition under antecedent-model conditions. The results favouring antecedent model in this review may therefore be because the antecedent model generates higher levels of correct responding during practice than error-correction procedures.

A major weakness of the 10 experiments which have been reviewed is that 9 of the 10 studies did not control the number of response opportunities during practice. Participants probably responded more in the error-correction procedures as they were often provided with secondary response opportunities contingent upon errors. This was clearly the case in Hendrickson et al. (1978) and McCurdy et al. (1990). Only one of the non-controlled response-opportunity studies (Hendrickson et al., 1978) provided data which allowed the number of additional practice responses to be calculated. Although the error-correction treatment in Hendrickson et al. (1978) generated a greater number of practice responses than the antecedent-model treatment, it was not as effective. However, the data presented only shows information on practices during instruction. It is to be expected that fewer errors during instruction will occur under antecedent-model treatments. The one study where the error-correction procedure was more effective was Ellis et al. (1978). The fading procedure in the antecedent-model treatment required the participant to respond to a set number of trials. This number was higher than the number of trials that participants required in the error-correction treatment. So, while the participant may have responded correctly more often during instruction, the total number of practice responses was greater.

After reporting the results of four reviews and reviewing 10 experiments it is still unclear which is more effective with classroom tasks such as spelling: some form of antecedent model or some form of error-correction procedure. Although the results of the present review suggest that antecedent-modelling procedures are more effective, variation in the number of uncontrolled response opportunities occurred in 9 of the 10 studies reviewed. This raises the question of what would happen if the number of response opportunities was held constant.

Aim

The aim of Experiment 3 was to measure the effect on rate of acquisition and instructional efficiency of (a) presenting a spelling model prior to practice responses or (b) presenting that model following incorrect spelling responses with the number of response opportunities controlled.

Procedural Changes

The measurement procedure implemented for Experiments 1 and 2 employed a 24-hour, 48-hour and 72-hour probe test to measure acquisition. It was found that a word correct on the 24-hour probe test was almost always correct on the 72-hour probe test. In addition, the measurement procedure used in Experiments 1 and 2 had a number of disadvantages.

First, a word correct on the 24- or 48-hour probe test could be incorrect on a subsequent probe test. When this happened the incorrect word was returned to the practice set. This created a potential confound in that 24- or 48-hours had passed without practice on the words returned to the practice set. It is possible that this delay could affect rate of acquisition.

Second, the testing procedure was somewhat cumbersome to administer. In Experiment 1 all words practised were measured. Participants were tested each day on the 10 words practised 24-hours earlier as well as words practised 48- and 72-hours earlier. This resulted in a lengthy daily probe test that was too long for 6-year old children. In order to decrease the administration time, Experiment 2 participants were only tested on the first 10 words in any treatment. However, while this decreased the probe test administration time, it also resulted in a ceiling effect as participants could only acquire 10 words. This ceiling effect prevented measurement of instructional efficiency. Because the measurement of instructional efficiency was one of the aims of these investigations this was a shortcoming.

In order to avoid these problems, it was decided to change the measurement procedure for Experiment 3 by returning to the procedure of testing all practised words. To make this possible, it was decided to use words correct on the 24-hour probe test as the measure of acquisition. This reduced the daily probe test to 10 words each day while at the same time eliminating the ceiling effect observed in Experiment 2. It also allowed instructional efficiency to be calculated over more than just 10 words.

Although better managed than in Experiment 1, task difficulty was still not well controlled in Experiment 2. Known difficult-to-learn words had been removed from both the Spelling 6 and the Spelling 8 word lists. However, this alone was not sufficient to satisfactorily control the task difficulty. It was therefore decided to count the number of letter pairs incorrect and the total number of letter pairs in each word in the pool of unknown words for each learner and to assign words with equal or similar number of incorrect letter pairs and total number of letter pairs to each treatment for each child.

The efficiency measure employed in Experiment 1 and Experiment 2 was the number of trials to criterion (the number of practice trials per response required in order to acquire that response). For example, a learner might have required eight practice trials on the word *table* in order to acquire the correct spelling of the word *table*. The instructional efficiency would therefore be eight trials to criterion for the word *table*. Mean trials to criterion was measured for each participant for each treatment by dividing the number of practice trials on the acquired words in a treatment by the number of acquired words in that treatment. For example, a participant may have acquired 10 words and may have had 50 trials on those 10 words. The mean trials to criterion therefore would be five trials per word acquired.

It was decided that another efficiency measure might also be appropriate in future experiments. This would be the total number of practice responses on both the words acquired and the words not acquired divided by the total number of spelling responses acquired. This efficiency measure, the mean number of practice responses required to acquire a spelling response, was labelled *practice responses to criterion.* For example, a learner might have acquired 10 spelling responses after 5 sessions with 20 practice responses in each session (100 total practice responses). The mean practice responses to criterion would therefore have been 10 practice responses per spelling response acquired. Employing the practice responses to criterion measure in addition to the trials to criterion measure in future experiments appeared to have several advantages. First, practice responses to criterion provides data on total practice response effort. Second, it might allow a teacher to predict the number of responses a child might acquire under certain conditions prior to instruction. Third, previous experiments have used this procedure (Barbetta, Heron et al., 1993; Barbetta & Heward, 1993). This has allowed the experimenter to calculate practice responses to criterion and to make efficiency comparisons across different experiments.

Because Experiments 1 and 2 had demonstrated that the computer programmes provided a high degree of experimental control, it was decided to reduce the length of Experiment 3 from 10 days (Experiment 1 and 2) to 5 days for two reasons. Previous research (Alvarado-Gomez & Belfiore, 2000; Barbetta, Heward, & Bradley, 1993; Grskovic & Belfiore, 1996) has demonstrated that an experiment lasting five days can be sufficient to provide a reliable measure of treatment effects on rate of acquisition at this age level. Further, this procedure removes the potential confound that existed in Experiments 1 and 2 where the Monday probe tests were actually 72 hours and not 24 hours after the Friday practice session.

METHOD

Participants and Setting

Eleven Year 2 children participated in the experiments. Table 18 describes the characteristics of the participants. The selection and screening procedures were the same as in Experiments 1 and 2. All the children who were screened were selected for the experiment.

The experiments were conducted at the same computer stations in the library within the same school as in Experiment 2. As in Experiment 2, screens were placed between computer stations and fixed chairs were provided.

Learning Tasks

Participants were individually tested on a pool of words selected from the LYC Spelling 6 and/or LYC Spelling 8 programme. The participants were tested on the Spelling 6 word list until 50 unknown words were found. However, 50 unknown words were not located for Tatum, Joel, Rod, Becky, Daryl, or John. These participants were tested on the Spelling 8 word list to locate 50 unknown words.

Pre-Experimental Procedures

The pre-experimental procedures were the same as Experiment 2. Participants were provided with practice on the typing programme until they could copy type 20 correct letters per minute. Participants practised the spelling programme until they could operate it independently. Practice on the typing and spelling programme took two to three sessions.

It was decided to count the number of letter pairs incorrect and the total number of letter pairs in each word in the pool of unknown words for each child. To ensure that each set of words was of similar difficulty, words incorrect on the pretest were callibrated using a letter-pair callibration procedure described by White and Haring (1980). The experimenter counted the number of letter pairs incorrect and the total number of letter pairs in each word

Table 18

Characteristics of Participants in Experiment 3

Participant	Gender	Age (years, months)	Reading level[a]	Spelling programme employed
1 Jeff	Male	6.6	15 (6.5 years)	Spelling 6
2 Teona	Female	6.5	17 (7.0 years)	Spelling 8
3 Tatum	Female	6.7	23 (8.0-8.5 years)	Spelling 8
4 Joel	Male	6.8	23 (8.0-8.5 Years)	Spelling 8
5 Catherine	Female	6.3	23 (8.0-8.5 years)	Spelling 6
6 Leigh	Female	6.6	17 (7.0 years)	Spelling 6
7 Rod	Male	6.10	22 (8.0-8.5 years)	Spelling 8
8 Leon	Male	6.9	17 (7.0 years)	Spelling 6
9 Reagan	Female	6.8	17 (7.0 years)	Spelling 6
10 Becky	Female	6.10	18 (7.0-7.5 years)	Spelling 8
11 Darryl	Male	6.8	23 (8.0-8.5 years)	Spelling 8
12 John	Male	6.5	21 (8.0 years)	Spelling 8

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

in the pool of unknown words. Pairs of incorrect words with the same number of unknown letter pairs were then alternately assigned to either Treatment A or B in an attempt to create two treatments of similar difficulty, that is, two treatments in which the total number of letter pairs to be learned and the total number of letter pairs correct on the pretest were closely matched. The same procedure that was used to construct the two sets of words was used to replace words spelled correctly on the 24-hour probe test for each treatment. This ensured that the difficulty level of the sets remained equivalent throughout the experiment.

Measurement Procedures

Twenty four hours after each practice session (and prior to the next day's practice session) the experimenter individually tested each participant on a 24-hour probe test using the same oral testing procedure which had been used in Experiments 1 and 2. Each word spelled correctly on the 24-hour probe test was removed from the practice set and replaced with a word from the pool of unknown words for that child. Additional 48-hour and 72-hour probe tests were not administered.

General Procedure

The general procedure was the same as that used in Experiments 1 and 2 except that each treatment lasted four days beginning on Monday and concluding on Thursday. Twenty four hour probe testing occurred each day for four days beginning on a Tuesday and concluding on a Friday. Prior to the first session of the experiment the experimenter loaded the first 10 words for each treatment for each participant in the computer using the "Set Editor" as described in Chapter 3. Prior to each subsequent session the experimenter removed words correct on the probe test from the practice set and replaced them with words from the pool of unknown words using the matching procedure described above. The computer functions were the same as those described in the General Procedure for all Experiments Section in Chapter 3.

Every practice session contained 30 trials. Each day's word set contained 10 words which were practised three times. After the 10 words had been presented to the participant on the first round the computer re-presented the first word again and re-presented words two to ten in a new random order. The computer then re-presented the list of words a third time using this same procedure. The spelling session was completed when the participant pressed the RETURN key after the thirtieth trial. Apart from the first session, Tatum practised for two rounds in her sessions as she found three rounds too long and became very fatigued.

Computer programming modifications. In Experiments 1 and 2 the intertrial interval was 0 seconds following a correct response and 8 seconds following an error response. In Experiment 3 the RIGHT ARROW key was activated following the participant's response and performed the function of taking the participant to the next trial. This gave participants control over the intertrial interval and allowed them the opportunity to spend as much time as they wanted to attend to each error-contingent model.

Antecedent-Model Treatment. For all trials in the Antecedent-Model Treatment the computer presented the sentence and a model of the word by displaying the word above the sentence (and below the picture for Spelling 6) as shown in Figure 28. At the same time a

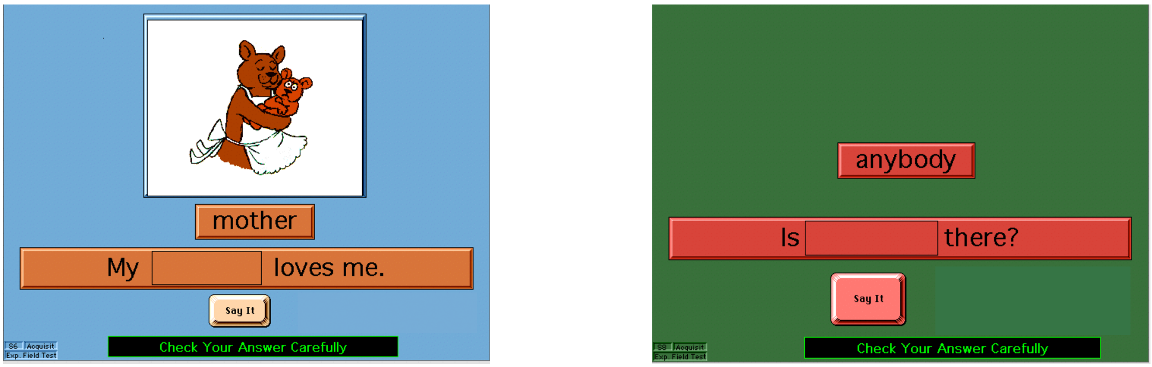


Figure 28. Antecedent-model screens for Spelling 6 and Spelling 8 programmes for the Antecedent-Model Treatment in Experiment 3.

computer-generated voice said the sentence. Once the participant began to type their response the computer removed the model of the target word from the screen. The participant completed their response by pressing RETURN. If the participant responded correctly the computer (a) provided feedback in the form of a ✔ at the end of the sentence as shown in Figure 29 and (b) presented the correct answer sound. If the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at the end of the sentence as shown in Figure 30 and (b) the incorrect answer sound. The participant pressed the RIGHT ARROW key to move to the next trial.

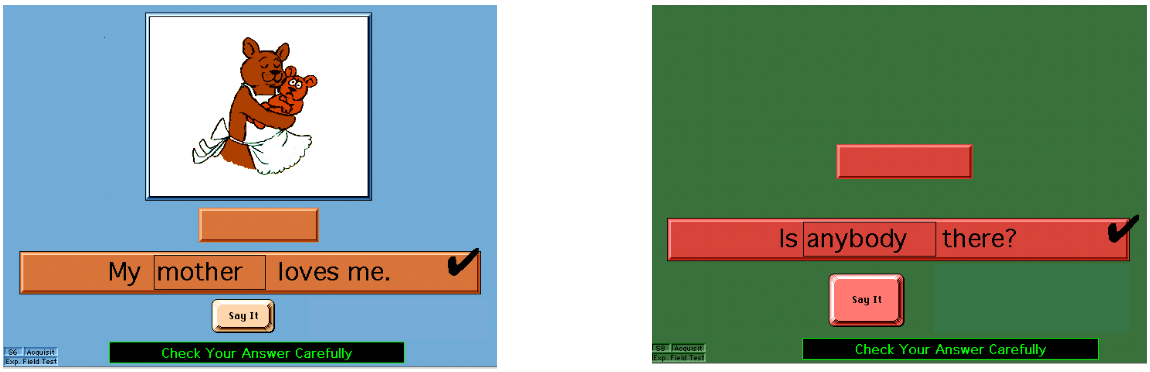


Figure 29. S6 and S8 completed response screens with correct feedback for the Antecedent-Model Treatment in Experiment 3.

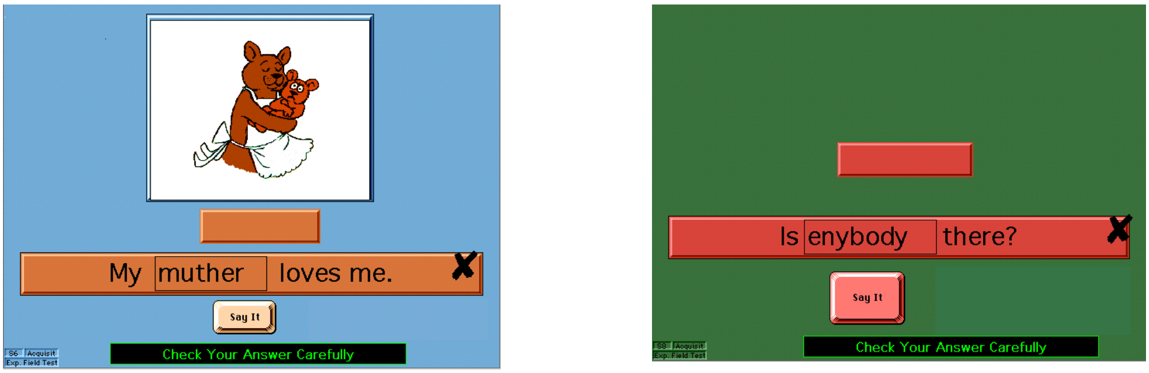


Figure 30. S6 and S8 completed response screens with error feedback for the Antecedent-Model Treatment in Experiment 3.

Error-Contingent Model Treatment. Prior to each error-contingent model session the experimenter reminded each participant that the *clue* (the model of the correct spelling) came after they had a go at the word and so it was important to have a go even if they didn't know how to spell the word. All trials in the Error-Contingent Model Treatment began with the computer (a) presenting the sentence as shown in Figure 31 and (b) presenting the computer-generated voice saying the sentence. No model of the target word was presented. If the participant responded correctly the computer provided (a) feedback in the form of a ✔ at the end of the sentence as shown in Figure 22 (Chapter 5) and (b) the correct answer sound. If the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at

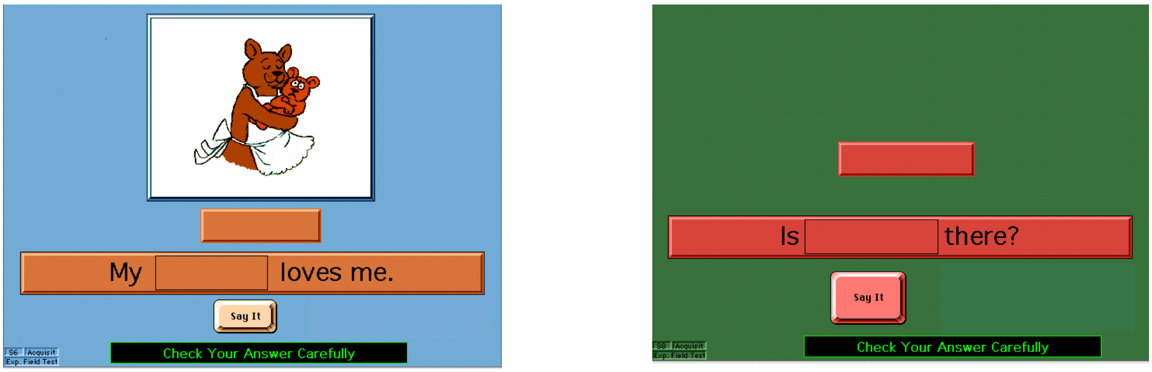


Figure 31. S6 and S8 trial screens (prior to responding) for the Error-Contingent Model Treatment in Experiment 3.

the end of the sentence and a model of the word above the sentence as shown in Figure 32 and (b) the incorrect answer sound. When ready, the participant pressed the RIGHT ARROW key to move to the next trial.

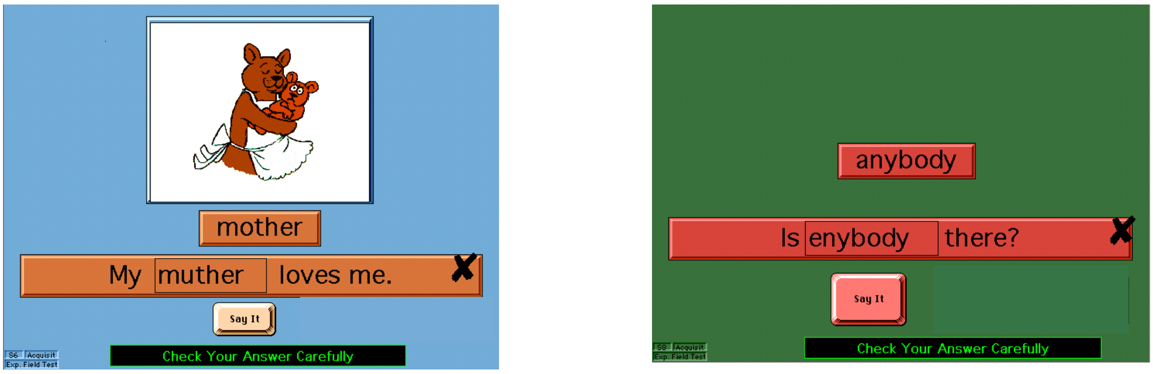


Figure 32. S6 and S8 completed response screens with error feedback and error-contingent model for the Error-Contingent Model Treatment in Experiment 3.

Experimental Design

An alternating treatments design (Cooper, Heron, & Heward, 1987) was used. The participants attended two sessions per school day; an Antecedent-Model Treatment and an Error-Contingent Model Treatment. One session was at 9.00 a.m. and the other at 1.30 p.m. The order of each treatment was arranged so that each participant experienced each treatment twice in the morning and twice in the afternoon. All participants were run individually. The experiment was replicated across 12 children.

RESULTS

Interscorer Agreement

A second-year teacher trainee conducted accuracy checks on the scoring of 25% of participants' responses on the probe tests. Interscorer agreement was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying this by 100. Interscorer agreement was 100%.

Procedural Reliability

A second-year teacher trainee conducted procedural reliability checks on 25% of the sessions to ensure that the participant correctly received either the Antecedent-Model Treatment or the Error-Contingent Model Treatment. Procedural reliability was assessed on agreements and disagreements between the within-session treatment printout and the experimenter's recording sheet of a session within each treatment. The procedural reliability was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying by 100. Procedural reliability was 100%.

Treatment Integrity

Treatment integrity was assessed by viewing the computer printouts showing the participants' responses and the position of the model on all words practised. It was found that the independent variable in both treatments was implemented as planned for 11 of the 12 participants. It was discovered that the computer did not maintain the set parameters under the Error-Contingent Model Treatment for Participant 2, Teona, on the first three days of the experiment. This meant that the set size and the number of practice responses per word varied from that experienced by the other 11 children. On Day 1, only nine words were presented in the first error-contingent model session. The word *paint* was practised 10 times and the word *smallest* was practised only once. The reason for this malfunction is unknown. Treatment integrity was therefore achieved for 11 of the 12 participants. Because of the computer malfunction, Teona's results were withdrawn from the experiment.

Rate of Acquisition

The cumulative number of words correct on the four 24-hour probe tests for each participant under each treatment is shown in Figure 33. Words correct on the 24-hour probe test were classified as acquired.

Effectiveness

Figure 33 shows that the Antecedent-Model Treatment was more effective for five of eleven participants. For three of these participants the separation of data points favouring the Antecedent-Model Treatment occurred on the first probe test. The Error-Contingent Model Treatment was more effective for one participant (Leon) and there was little or no difference between the effect of the two treatments for four participants. Jeff's data was incomplete as he was absent for the final session for the Error-Contingent Model Treatment. However, error-contingent modelling was more effective after three probe tests. As Jeff only completed three sessions in the Error-Contingent Model Treatment, his data was not included in calculating the mean effectiveness across treatments.

Table 19 shows that the number of words acquired ranged from 6 words to 21 words in the Antecedent-Model Treatment, and from 7 words to 22 words in the Error-Contingent Model Treatment. The mean number of words acquired was 12.6 words for the Antecedent-Model Treatment, and 11.3 words for the Error-Contingent Model Treatment. The difference was not significant ($t = 1.4$, $p < .05$).

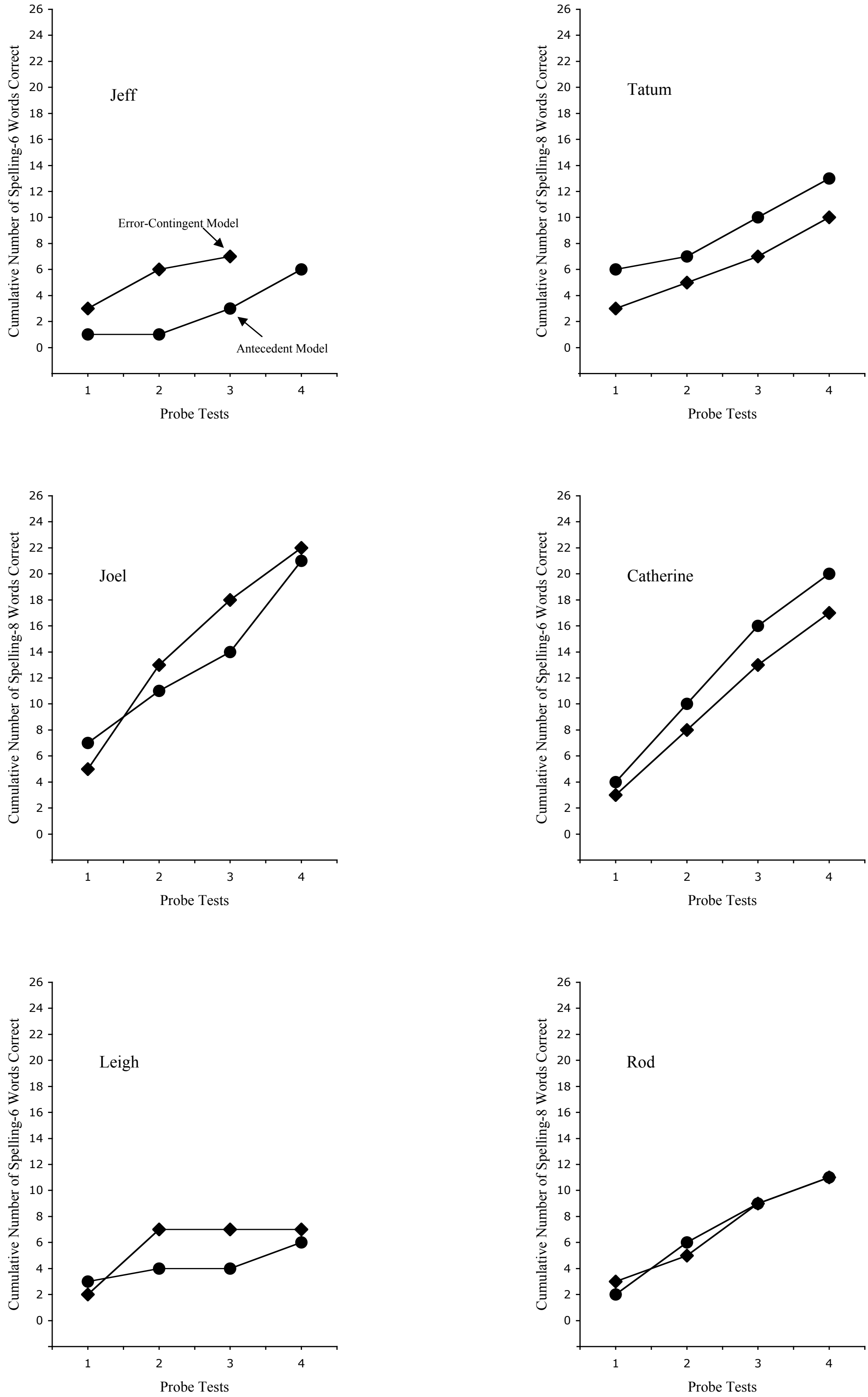


Figure 33. Cumulative number of words correct on the 24-hour probe test for each participant under each treatment.

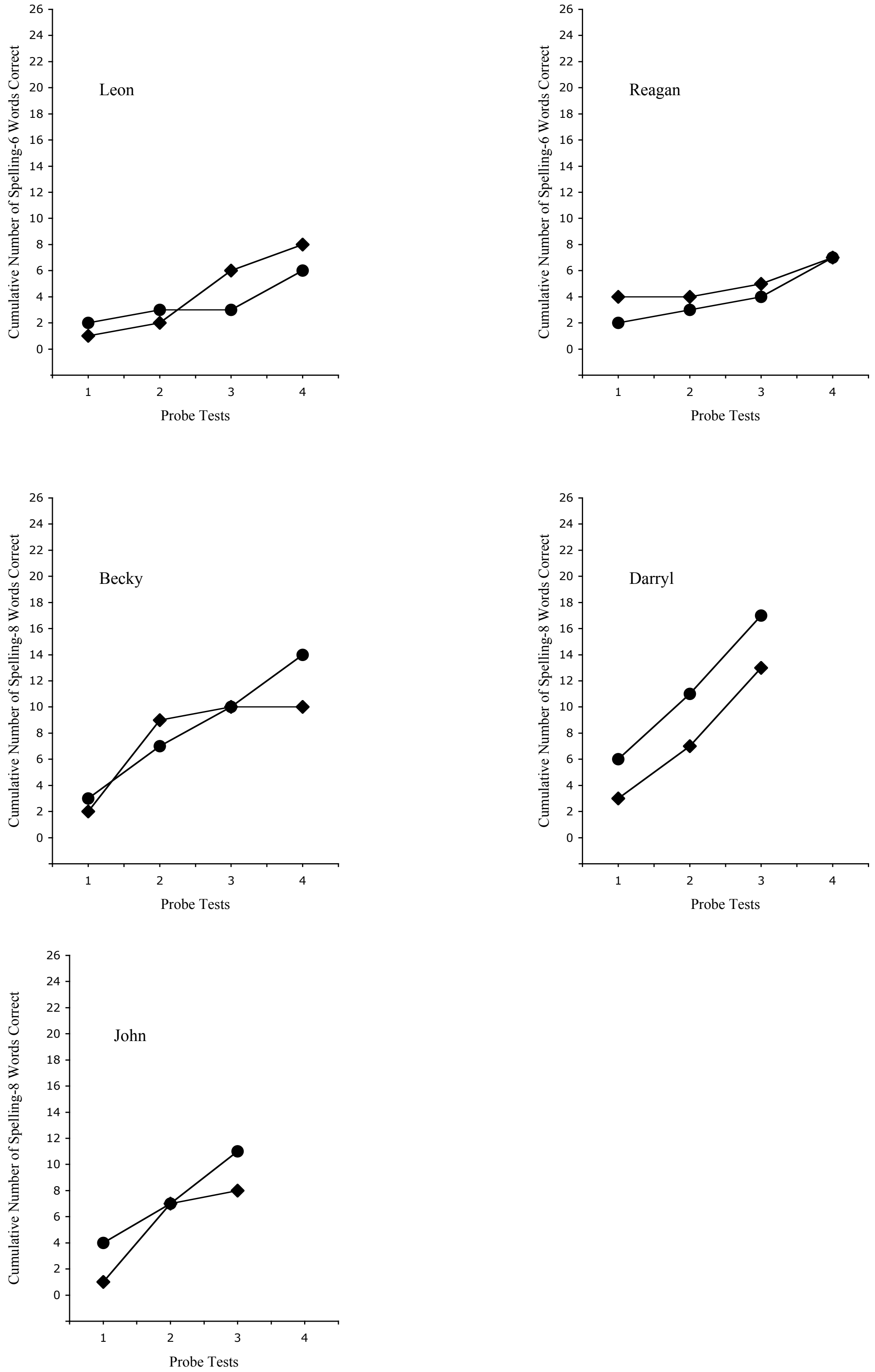


Figure 33. (Continued).

Table 19

Effectiveness, Trials to Criterion, and Practice Responses to Criterion for Participants Under Antecedent-Model and Error-Contingent Model Treatments in Experiment 3

	Effectiveness (Number of words acquired after 4 sessions)		Trials to criterion (mean number of practice responses on acquired words per word acquired)		Practice responses to criterion (mean number of practice responses on all words per word acquired)	
	Antecedent Model	Error-Contingent Model	Antecedent Model	Error-Contingent Model	Antecedent Model	Error-Contingent Model
Jeff	-[a]	-	6.2	3.9	15.0	17.1
Tatum	13	10	6.8	7.2	6.9	9.0
Joel	21	22	5.1	4.1	5.7	5.5
Catherine	20	17	4.1	4.9	6.0	7.1
Leigh	6	7	6.0	4.7	20.0	17.1
Rod	11	11	5.7	6.3	10.9	10.9
Leon	6	8	5.5	7.0	18.3	13.8
Reagan	7	7	7.7	5.6	17.1	17.1
Becky	14	10	4.4	5.2	8.6	12.0
Darryl	17	13	4.4	4.5	7.1	9.2
John	11	8	4.5	5.6	10.0	13.8
Mean	12.6	11.3	5.5	5.4	11.4	12.1
SD	5.5	4.9	1.1	1.1	5.3	4.1

[a] As Jeff's data was incomplete his scores were not used when calculating the mean effectiveness after four sessions. Effectiveness for Jeff was 6 words acquired after four sessions for the Antecedent-Model Treatment and 7 words acquired after three sessions for Error-Contingent Model Treatment.

Trials to Criterion

The mean number of trials to criterion ranged from 4.1 trials for Catherine to 7.7 trials for Reagan in the Antecedent-Model Treatment, and from 3.9 trials for Jeff to 7.2 trials for

Tatum in the Error-Contingent Model Treatment. Across participants the mean number of trials to criterion was 5.5 for the Antecedent-Model Treatment and 5.4 for the Error-Contingent Model Treatment. The difference was not significant ($t = 0.7$, $p < .05$).

Practice Responses to Criterion

The mean total number of practice responses required ranged from 5.7 practice responses per word acquired for Joel to 20.0 practice responses per word acquired for Leigh in the Antecedent-Model Treatment, and from 5.5 practice responses per word acquired for Joel to 17.1 practice responses per word acquired for Jeff, Leigh and Reagan in the Error-Contingent Model Treatment. Across participants the mean practice responses to criterion was 11.4 practice responses per word acquired for the Antecedent-Model Treatment and 12.1 practice responses per word acquired for the Error-Contingent Model Treatment. The difference was not significant ($t = 0.8$, $p < .05$).

Participant's Responses to Instruction

In order to evaluate the social validity of the experiments the experimenter interviewed each participant at the end of the experiment. To determine which treatment the participant preferred the experimenter asked, "How did you like to learn your spelling best – with the clue before you had a go at the word or with the clue after you had a go at the word?" To determine how difficult the learner found each experimental treatment the experimenter asked each participant, "When you were learning your spelling words with the clue before (or after) you had a go at the word, did you find the work on the computer easy, middle or hard?" To determine whether the participant enjoyed working on the computer in each treatment the experimenter asked, "When you were working on the computer and you had the clue before (or after) you had a go at the word did you not enjoy it, enjoy it a bit, or enjoy it a lot." Social validity data is presented in Table 20. Becky was the only participant who stated that she "liked both (treatments) the same." John was absent on the day of the

Table 20

Participants' Social Validity Reports for Experiment 3

	Enjoyment		Difficulty level		Treatment preference
	Antecedent model	Error-Contingent Model	Antecedent Model	Error-Contingent Model	
Jeff	Did not enjoy	A bit	Easy	Hard	Antecedent Model
Tatum	A lot	A lot	Easy	Easy	Antecedent Model
Joel	A lot	A bit	Easy	Easy	Error-Contingent Model
Catherine	A bit	A bit	Easy	Easy	Antecedent Model
Leigh	A lot	A lot	Easy	Easy	Error-Contingent Model
Rod	A lot	A lot	Easy	Easy	Antecedent Model
Leon	A bit	A lot	Easy	Easy	Antecedent Model
Reagan	A lot	A lot	Easy	Easy	Antecedent Model
Becky	A bit	A bit	Middle	Easy	Both the same
Darryl	A lot	Did not enjoy	Easy	Hard	Antecedent Model
John [a]	-	-	-	-	-

[a] Indicates that no social validity data was collected

final probe test so no data is reported on experimental preference at the conclusion of his experiment. He was however asked which treatment he preferred twice during the experiment (Days 1 and 3). On both occasions he reported liking the Antecedent-Model Treatment best.

Jeff. Jeff acquired six spelling responses after four sessions for the Antecedent-Model Treatment. He was absent for one session in the Error-Contingent Model Treatment and acquired seven spelling responses after three sessions. Given the rates of acquisition of each treatment it is difficult to say which treatment would have been more effective after four sessions. Jeff reported that he enjoyed the Error-Contingent Model Treatment "a bit" and "didn't enjoy" the Antecedent-Model Treatment, yet stated that his treatment preference was the Antecedent-Model Treatment. About half way through the first round of an Error-Contingent Model session, Jeff stated, "I'm getting them all wrong."

Tatum. Tatum's rates of acquisition were parallel except for the first session where she learned three more words in the Antecedent-Model Treatment. The mean number of trials to criterion (mean number of practice responses on acquired words per word acquired) was the same (7) for each treatment. Tatum said that she enjoyed both treatments "a lot" and found them both "easy."

Joel. Joel's rates of acquisition were the same for both treatments as he acquired just over 20 spelling responses in both conditions. He reported preferring the Error-Contingent Model Treatment and stated that he found both treatments "easy." At the beginning of the second round on one session of the Antecedent-Model Treatment Joel stated, "I am getting them all right."

Catherine. Catherine acquired 20 spelling responses and 17 spelling responses in the Antecedent-Model Treatment and Error-Contingent Model, respectively. She reported that she preferred the Antecedent-Model Treatment and found both treatments easy. During the first round of an Error-Contingent Model session Catherine stated, "I haven't got any right yet."

Leigh. Rates of acquisition and trials to criterion were almost identical in both treatments for Leigh. However, Leigh failed to acquire any spelling responses in the Error-Contingent Model Treatment after the second session, and didn't acquire any spelling responses in the third session in the Antecedent-Model Treatment. Leigh reported that both treatments were "easy" and that she preferred the Error-Contingent Model Treatment. It seems the decrease in the rates of acquisition for both treatments was because Leigh attended less to the model as the experiment continued. Her accuracy levels in the Error-Contingent Model Treatment were 14% and 28% respectively for the first two sessions decreasing to 0% and 7% respectively for the last two sessions. This trend was also observed for the Antecedent-Model Treatment.

Rod. There was no difference in rates of acquisition or trials to criterion between the two treatments for Rod. He reported enjoying both treatments “a lot” and found both “easy.”

Leon. Leon rates of acquisition for both treatments were parallel except for the third session when they crossed over as he acquired no spelling responses in the Antecedent-Model Treatment. It was unlikely this was due to the accuracy level in the Antecedent-Model Treatment as it was comparable to other sessions suggesting that attending to the model was high. An inspection of the words in both treatments showed that word difficulty was well controlled in that equivalent letter-pairs were unknown. However, Leon never acquired six of the first 10 words practised suggesting that word difficulty may have in fact been higher in the Antecedent-Model Treatment. Leon reported both treatments as “easy” and stated that he enjoyed the Error-Contingent Model Treatment “a lot” and the Antecedent-Model Treatment “a bit,” yet described this as his preferred treatment.

Reagan. Reagan’s rates of acquisition were parallel as she acquired seven spelling responses in each of the two treatments. She reported enjoying both treatments “a lot” and said they were both “easy.”

Becky. Becky acquired 14 spelling responses in the Antecedent-Model Treatment and 10 spelling responses in the Error-Contingent Model Treatment. The rates of acquisition were essentially parallel after the first two sessions however Becky only acquired one spelling response in the last two sessions of the Error-Contingent Model Treatment. Becky described the Error-Contingent Model Treatment as “easy” yet during a session in this treatment was observed to say to herself, “I don’t enjoy things that are hard.” As the accuracy levels during the last two sessions of the Error-Contingent Model Treatments were 7% and 0% respectively it seems that Becky found this treatment aversive and did not attend to the error-contingent model as a result.

Darryl. Apart from the first session, Darryl's rates of acquisition within each treatment were parallel and his trials to criterion were equal. While words were equally matched for unknown letter-pairs at the beginning of the experiment, Darryl did not acquire two of the first 10 words practised in the Error-Contingent Model Treatment. This suggests that these two words were more difficult than corresponding words in the Antecedent-Model Treatment. On the word *paint* (a response which was acquired) in the Error-Contingent Model Treatment, Darryl stated that he didn't know the word. When prompted by the experimenter to attempt it, he stated that he would get it wrong. The experimenter asked if he liked getting words wrong. Darryl replied, "No." Darryl reported that he found the Error-Contingent Model Treatment "hard" and that he "did not enjoy" it. Darryl was a high-achieving student who did not like making errors and found this treatment aversive. This was likely because the error-contingent model did not control the correct practice response to the same level as the antecedent model. Darryl therefore had a lower accuracy level during instruction than he was probably used to in the classroom.

John. John acquired 11 spelling responses in the Antecedent-Model Treatment and 8 spelling responses in the Error-Contingent Model Treatment after three sessions. John only acquired one word in the final session of the Error-Contingent Model Treatment. This was likely due to two factors. First, three of the first 10 words practised were not acquired. Although both word sets were equally matched for difficulty (three of the first 10 words practised in the Antecedent-Model Treatment were also not acquired), the words were too difficult. Second, John had acquired six words the previous session which meant that six new words were added to the practice set. As previously discussed, it seems that the children required about five trials in order to acquire a spelling response. As these six words were only practised three times it appears that there was insufficient practice.

Accuracy Levels During Instruction

The accuracy levels during instruction for each participant across treatments for words acquired, non-acquired, and all words practised can be seen in Table 21. Accuracy levels during instruction ranged from 57% for Leigh to 92% for Darryl in the Antecedent-Model Treatment, and from 8% for Reagan to 50% for Darryl in the Error-Contingent Model Treatment. The mean accuracy level during instruction across participants was 76% for the Antecedent-Model Treatment and 25% for the Error-Contingent Model Treatment. In all cases, the accuracy level during instruction was higher for acquired words than for non-acquired words for each participant within each treatment.

Effects of the Models

To calculate the percentage of responses correct following the antecedent model for the first round of a session, the number of models and the number of responses correct following each antecedent model were counted for the first 10 responses per session for all antecedent-model sessions. In the Error-Contingent Model Treatment the percentage of responses correct following the error-contingent model for the first round of a session was calculated by counting the number of words correct on the second round of a session which had been incorrect (and had therefore received an error-contingent model) on the first round of a session. This was converted to the percentage of responses correct following the presentation of an error-contingent model during the first round of a session. The results of this analysis are shown in Table 22. The effectiveness of the antecedent model in the first round ranged from 55% for Tatum, Leigh and Reagan to 93% for Darryl. The mean percentage of correct responses following the antecedent model for the first round of a

Table 21

Accuracy Levels During Instruction for Antecedent-Model and Error-Contingent Model Treatments for Each Participant in Experiment 3

	Accuracy level during instruction					
Name	Antecedent Model			Error-Contingent Model		
	Acquired words	Non-acquired words	All words	Acquired words	Non-acquired words	All words
Jeff	92	91	91	59	3	20
Tatum	80	49	66	38	0	29
Joel	77	42	73	34	0	26
Catherine	86	67	80	54	12	42
Leigh	78	47	57	33	1	11
Rod	84	75	80	33	10	23
Leon	97	65	75	21	20	21
Reagan	87	34	59	23	0	8
Becky	85	67	78	33	9	25
Darryl	93	87	92	49	14	50
John	92	71	83	33	19	25
Mean	86	63	76	37	8	25

session was 70%. It can be seen that the mean percentage of correct responses following the model was higher for acquired words (81%) than non-acquired words (56%).The effectiveness of the error-contingent model (the consequent model) in the first round ranged from 5% for Reagan and Becky to 56% for Leigh. The mean percentage of correct responses following the error-contingent model for the first round of a session was 23%. The mean percentage of correct responses following the error-contingent model was higher for acquired words (26%) than non-acquired words (9%).

Table 22

Mean Percentage of Words Presented for the First Time Which Were Correct Following a Model in (a) the Antecedent and (b) the Consequent Position in Experiment 3

Name	Antecedent Model			Error-Contingent Model		
	Acquired words	Non-acquired words	All words	Acquired words	Non-acquired words	All words
Jeff	85	83	84	20	5	8
Tatum	68	39	55	32	0	15
Joel	78	50	75	33	0	24
Catherine	74	31	60	59	27	48
Leigh	83	42	55	11	4	56
Rod	81	65	74	17	12	14
Leon	92	59	69	16	13	14
Reagan	72	40	55	15	0	5
Becky	77	63	71	13	0	5
Darryl	96	80	93	40	14	32
John	88	69	80	33	27	30
Mean	81	56	70	26	9	23

DISCUSSION

In Experiment 3 it was found that 5 out of 11 of the 6-year old children acquired new spelling responses as quickly without an antecedent model as with an antecedent model.

Based on the findings of ten previous experiments, it was expected that the Antecedent-Model Treatment would be more effective and more instructionally efficient than the Error-Contingent Model Treatment. It was thought that this would happen because placement of the prompt in the antecedent position would have tighter stimulus control over the learner's spelling response and hence would generate a higher proportion of correct responses during practice. The findings of Experiment 2 further suggested that teaching procedures which produced a higher proportion of correct responses during practice also produced a higher rate of acquisition. The greater the number of responses correct during practice in Experiment 2 the greater the number correct on the 24-hour probe test. Although the proportion of correct responses during practice was higher in the Antecedent-Model Treatment than in the Error-Contingent Model Treatment in Experiment 3, there was only a slight difference in the number of words acquired and no difference in the responses to criterion. These results were quite unexpected and raise the question of why a teaching procedure which produced a lower proportion of correct responses during practice was almost as effective and was as instructionally efficient as one which produced higher proportions of correct responses during practice. Several reasons are explored.

First, it is possible that participants attended to the model differently depending upon whether it was in the antecedent position or the consequent position. When in the antecedent position, participants would often attend to and self-rehearse all the letter pairs of the word whilst viewing the model of the word. It is likely that participants self-rehearsed all letter pairs because they did not know which letter pairs they would respond to correctly or incorrectly. This differed when the model was in the consequent position as participants could compare the model with their incorrect response. This allowed them to focus on the incorrect letter pairs of the word rather than all letter pairs in the word. This may have decreased the number of letter pairs selected for self-rehearsal under this treatment when compared with the number of letter pairs selected for self-rehearsal under the antecedent-model treatment.

Second, it may be that participants have experienced models in the consequent position more often than in the antecedent position. Apart from the daily reading session with the teacher, the participants may not have had experience with highly prompted tasks within the classroom. That is, participants may have been more familiar with the error-contingent model procedure as it reflected classroom practice more closely than the antecedent-model procedure.

Third, the 24-hour probe testing procedure may have favoured the Error-Contingent Model Treatment in that no models of the correct spelling were presented with the practice stimulus during the probe test. That is, the probe testing procedure more closely matched the Error-Contingent Model Treatment than the Antecedent-Model Treatment. Donahoe and Palmer (1994, p. 231) suggest that retention was "better when the environment during the test was similar to the study environment in which the target response was strengthened."

It is clear from Table 22 that in the present experiment the antecedent model gained stimulus control over the participant's spelling response more often during instruction than did the model in the consequent position. Seventy percent of responses following the antecedent model were correct whereas only 23% of second-round responses following the error-contingent model were correct. This is to be expected given the simultaneous presentation of the practice stimulus and the presentation of the antecedent model whereas the latency between the presentation of the error-contingent model and the presentation of the practice stimulus was anywhere from approximately 30 seconds to 14 minutes. In addition, no other stimuli were presented in the period between the presentation of the antecedent model and the start of the participant's spelling response. This was not the case following the error-contingent model as other words were practised during the period between the presentation of this model and the next presentation of the stimulus question for that word.

In the Antecedent-Model Treatment, as soon as the participant began to type their response, the model disappeared from the screen. A correct practice response was a learning opportunity under the Antecedent-Model Treatment because the conditions existed for a transfer of stimulus control from the model of the correct word to the practice stimulus (the spoken word). Under the Antecedent-Model Treatment, errors were not learning opportunities because no correction (no error-contingent model) was presented and hence there was no opportunity for a transfer of stimulus control to occur following an error. However, in the Error-Contingent Model Treatment errors were learning opportunities because the presentation of the error-contingent model (the correction) provided an opportunity for the transfer of stimulus control to occur.

Error Consequences as Aversive Stimuli

Seven of the eleven participants stated that they preferred the Antecedent-Model Treatment. During the Antecedent-Model Treatment participants would often call out that they had, for example, “got the last five right” (Leigh) or that they hadn’t “made a mistake yet” (Lucy), or “ I have only got one wrong” (Joel). It may be that some participants acquired the words under a (predominantly) positive reinforcement contingency in this treatment. For example, a participant was presented with the stimulus question and the model, responded correctly, and was then presented with positive feedback in the form of a ✔. The presentation of the tick may have functioned as a positive reinforcer. Participants also received high rates of positive feedback under the Antecedent-Model Treatment (76% of responses) compared to the Error-Contingent Model Treatment (25% of responses).

In the Error-Contingent Model Treatment some participants complained during the first round of a session (where it would be expected that the accuracy level would be 0%) that they were “getting them all wrong” (Jeff) or “I haven’t got any right yet” (Catherine). It may be that participants acquired the words in the Error-Contingent Model Treatment predominantly under a negative reinforcement contingency. A participant was presented with the stimulus question, they responded incorrectly, received negative feedback and a model of the correct response which they then attended to in order to avoid an error on the next oral presentation of the word. However, the present computer set up also allowed an escape response. The participant could escape the negative feedback and error-contingent model simply by pressing the RIGHT ARROW key (this resulted in the immediate presentation of the next trial). It appears both of these contingencies operated for different words within particular sessions. For example, Leigh attended closely to the error-contingent model on words such as *took* and *put* (acquired words) and yet was observed to immediately press the RIGHT ARROW key whenever the error-contingent model was presented on words such as *night* and *could* (non-acquired words).

Stimulus Control Provided by the Antecedent Model

As in Experiments 1 and 2, the antecedent model did not always function as a strong prompt in Experiment 3. The antecedent model functioned to control correct responding 76% of the time. The reasons for the failure of the model to control correct responding on 24% of occasions are probably the same as those described in Chapter 6. Namely that the model was a non-copying prompt, some words may have been too difficult to reproduce correctly on the first trial under these conditions, and the attending behaviours of the participants tended to be somewhat variable.

Task Difficulty

The task difficulty was considerably better controlled in Experiment 3 than in Experiment 2. This was because words were pretested for letter-pair errors, matched for difficulty and assigned pair-wise to the two treatments.

Implications

Prompting and fading procedures take more time to set up than error-correction procedures in the classroom (Etzel & LeBlanc, 1979). Given that the antecedent-model procedure was only slightly more effective and slightly more instructionally efficient, it might be argued that it does not matter too much which procedure a teacher uses in the classroom. However, the higher number of errors during instruction under error-contingent model conditions and the possibility that acquisition may be a function of negative rather than positive reinforcement needs to be taken into account otherwise learners may become discouraged and begin to dislike the task. This occurred for two children (Jeff and Darryl) in the Error-Contingent Model Treatment in the present experiment – which lasted less than two hours.

Summary

The results of Experiment 3 showed that there was little difference in the effectiveness, the mean number of trials to criterion, and the mean number of practice responses to criterion between an antecedent-prompting condition and an error-correction condition when the number of learning opportunities was controlled. Given that participants can learn new spelling words almost as well when the model is presented in the consequent position as when the model is presented in the antecedent position, it becomes clear that the way in which errors are handled is a matter of considerable importance in instructional situations such as the classroom where effective prompting procedures are often not possible.

REFERENCES

Aiken, E. G., & Lau, A. W. (1967). Response prompting and response confirmation. *Psychological Bulletin, 68*, 330-341.

Alvarado-Gomez, M., & Belfiore, P. J. (2000). Effects of error correction on Spanish spelling words for students whose primary language is Spanish. *Journal of Behavioral Education, 10*, 213-221.

Ault, M. J., Wolery, M., Doyle, P. M., & Gast, D. L. (1989). Review of comparative studies in the instruction of students with moderate and severe handicaps. *Exceptional Children, 55*, 346-356.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition and maintenance of geography facts by elementary students with learning disabilities. *Journal of Behavioral Education, 3*, 217-233.

Barbetta, P. M., Heward, W. L., & Bradley, D. M. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110.

Bennett, K., & Cavanaugh, R. A. (1998). Effects of immediate self-correction, delayed self-correction, and no correction on the acquisition and maintenance of multiplication facts by a fourth-grade student with learning disabilities. *Journal of Applied Behavior Analysis, 31*, 303-306.

Browder, D. M., & Lalli, J. S. (1991). Review of research on sight word instruction. *Research in Developmental Disabilities, 12*, 203-228.

Browder, D. M., & Xin, Y. P. (1998). A meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education, 32*, 130-153.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus: Merrill Publishing Company.

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Massachusetts: Allyn and Bacon.

Drevno, G. E., Kimball, J. W., Possi, M. K., Heward, W. L., Gardner, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and generalization of science vocabulary by elementary students: A systematic replication. *Journal of Applied Behavior Analysis, 27*, 179-180.

Duffy, L., & Wishart, J. G. (1987). A comparison of two procedures for teaching discrimination skills to Down's Syndrome and non-handicapped children. *British Journal of Psychology, 57*, 265-278.

Egeland, B. (1975). Effects of errorless training on teaching children to discriminate letters of the alphabet. *Journal of Applied Psychology, 60*, 533-536.

Ellis, W. D., Ludlow, B. L., & Walls, R. T. (1978). Learning, transfer, and retention of errorless fading versus trial-and-error teaching. *Psychological Reports, 43*, 553-554.

Espin, C. A., & Deno, S. L. (1989). The effects of modeling and prompting feedback strategies on sight word reading of students labeled learning disabled. *Education and Treatment of Children, 12*, 219-231.

Etzel, B. C., & LeBlanc, J. M. (1979). The simplest treatment alternative: The law of parsimony applied to choosing appropriate instructional control and errorless-learning procedures for the difficult-to-teach child. *Journal of Autism and Developmental Disorders, 9*, 361-382.

Everett, F. L. (1977). Role of errors in children's discrimination learning. *Perceptual and Motor Skills, 45*, 1187-1201.

Gast, D. L., Ault, M. J., Wolery, M., & Doyle, P. M. (1988). Comparison of constant time delay and the system of least prompts in teaching sight word reading to students with moderate retardation. *Education and Training in Mental Retardation, 23*, 117-128.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.

Gollin, E. S., & Savoy, P. (1968). Fading procedures and conditional discrimination in children. *Journal of the Experimental Analysis of Behavior, 11*, 443-451.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: HarperCollins.

Griffiths, K., & Griffiths, R. (1976). Errorless establishment of letter discriminations with a stimulus fading procedure in pre-school children. *Perceptual and Motor Skills, 42*, 387-396.

Grskovic, J. A., & Belfiore, P. J. (1996). Improving the spelling performance of students with disabilities. *Journal of Behavioral Education, 6*, 343-354.

Hendrickson, J., Roberts, M., & Shores, R. E. (1978). Antecedent and contingent modeling to teach basic sight vocabulary to learning disabled children. *Journal of Learning Disabilities, 11*, 69-73.

Heubusch, J. D., & Lloyd, J. W. (1998). Corrective feedback in oral reading. *Journal of Behavioral Education*, 63-79.

Johnson, P., Schuster, J., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6*, 437-458.

Koscinski, S. T., & Gast, D. L. (1993). Use of constant time delay in teaching multiplication facts to students with learning disabilities. *Journal of Learning Disabilities, 26*, 533-544.

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559-578.

McCurdy, B. L., Cundari, L., & Lentz, F. E. (1990). Enhancing instructional efficiency: An examination of time delay and the opportunity to observe instruction. *Education and Treatment of Children, 13*, 226-238.

McGee, G. G., & McCoy, J. F. (1981). Training procedures for acquisition and retention of reading in retarded youth. *Applied Research in Mental Retardation, 2*, 263-276.

Nelly, E., & Smith, A. (2000). *PM Benchmark kit: An assessment resource for emergent - 12 years reading age*. Melbourne: Nelson ITP.

Robinson, P. W., & Storm, R. H. (1978). Effects of error and errorless discriminaton acquisition on reversal learning. *Journal of the Experimental Analysis of Behavior, 29*, 517-525.

Simonsen, F., & Gunter, L. (2001). Best practices in spelling instruction: A research summary. *Journal of Direct Instruction, 1*, 97-105.

Smeets, P. M. (1992). Prompt -aided instruction of mirror-image discriminations: Abrupt prompt removal, time delay, and self-evaluation. *Journal of Behavioral Education, 2*, 87-104.

Stevens, K. B., & Schuster, J. W. (1987). Effects of a constant time delay procedure on the written spelling performance of a learning disabled student. *Learning Disability Quarterly, 10*, 9-16.

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities, 33*, 114-136.

Trap, J. J., Milner-Davis, P., Joseph, S., & Cooper, J. O. (1978). The effects of feedback and consequences on transitional cursive letter formation. *Journal of Applied Behavior Analysis, 11*, 381-393.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching (2nd ed.)*. Ohio: Charles E. Merrill Publishing Company.

CHAPTER 8

THE EFFECTS OF ERROR-CONTINGENT SECONDARY-RESPONSE OPPORTUNITIES

On typical classroom tasks, learners frequently produce errors. Even when learners were provided with an antecedent model of the current spelling response in Experiments 2 and 3, they still produced errors on approximately 30% of their practice responses. This raises the question of how errors should be handled in the classroom. The accepted view is that children's errors should be corrected during practice (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kulhavy, Yekovich, & Dyer, 1976). Errors may be corrected using a variety of error-correction procedures including self-correction, peer-correction or teacher-administered error correction (e.g., Skinner, Shapiro, Turco, Cole, & Brown, 1992), immediate or delayed error correction (e.g., Singh, Winton, & Singh, 1985), providing a weak or a strong prompt with the error correction (e.g., Espin & Deno, 1989), and/or providing a secondary response opportunity following the error (e.g., Barbetta, Heron, & Heward, 1993). An interesting question is whether it is more effective, following an error and an error correction, for the learner to practise the response again immediately, or whether it is sufficient to practise the response again at some later time.

One error-correction procedure is the so called active-student-response (ASR) error-correction procedure. This procedure involves each error correction trial ending "with the student emitting the correct response following a teacher-provided model" (Barbetta et al., 1993, p. 112). However, ASR error-correction is not a suitable term for a secondary-response treatment in this experiment for two reasons. First, "calling a response active is like saying water is wet" (Heward, 1994, p. 286). Second, Barbetta et al.'s definition assumes that the student will emit the correct response on the second trial. In the present experiments, participants have, on average, only responded correctly about 70% of the time following the first presentation of an antecedent model. Therefore, for the present experiment the term error-contingent model and secondary-response will be used to refer to the presentation of an error-contingent model of a word followed by a secondary response opportunity.

There are several possible advantages of the error-contingent model and secondary-response procedure over the error-contingent model procedure without a secondary response. First, the results of the review in Chapter 6 showed that error-correction procedures that provided for additional response opportunities were more effective than those that required no overt error-contingent response. Second, as the error-contingent model is presented immediately before the learner responds, transfer of stimulus control from the model to the spoken word may be more likely to occur than if the response is delayed and preceded by other practice trials. Third, informal observations of the children in Experiments 2 and 3 suggest that they often focused their attention on the specific letter pairs which they had typed incorrectly when the model was error contingent. Fourth, the provision of a secondary response opportunity may increase the accuracy level during instruction, increase the level of positive feedback and, possibly, reduce the level of aversiveness of the error-contingent feedback.

Two instructional efficiency measures have been employed in this thesis. In Experiment 3, practice responses to criterion was introduced to measure the mean number of practice responses on both the words acquired and the words not acquired in order to acquire a response. For example, a learner might have acquired 10 responses after 4 sessions containing 120 total practice responses. The mean practice responses to criterion would therefore have been 12 practice responses per response acquired. It will be recalled that trials to criterion measured the mean number of practice trials on acquired responses in order to acquire a response in Experiments 1, 2 and 3. For example, a learner might have required nine practice trials on the word *another* in order to acquire the correct spelling of this word.

The instructional efficiency would therefore have been nine trials to criterion for the word *another*. As participants had one response opportunity per trial in Experiments 1, 2 and 3 the number of responses was the same as the number of trials. However, in the present experiment it was planned to provide participants with a secondary-response opportunity within a trial. This meant that the number of responses would no longer be the same as the number of trials. This necessitated the introduction of a new measure of efficiency: responses to criterion. Responses to criterion was defined as the number of primary and secondary practice responses on an acquired response which were required in order to acquire that response.

The effect of providing secondary response opportunities following errors has been studied in a number of experiments. Of particular interest were experimental analyses in which (a) the participants were school-aged children, (b) the behaviours taught were discrete responses, and (c) the behaviours were academic, concept learning or discrimination responses. A literature search identified four such experiments (Barbetta et al., 1993; Barbetta & Heward, 1993; Drevno et al., 1994; Johnson, Schuster, & Bell, 1996). These are summarised in Table 23.

Characteristics of the experiments. All experiments employed an alternating-treatments design. Participants ranged in age from 8- to 17-years old. Two experiments (Barbetta & Heward, 1993; Johnson et al., 1996) classified the participants as having a learning disability and one (Barbetta et al., 1993) classified the participants as having a developmental disability. One experiment (Drevno et al., 1994) classified two participants as talented students and three as being at risk of academic failure. All experiments took place in a school setting. Interobserver reliability data was provided in all four experiments. The learning tasks are shown in Table 23.

Table 23

Results of Four Experimental Analyses of the Effects of ASR Error-Correction Procedures

Authors and date	Participants	Task	Measure	Design	Treatments	Timing of feedback	Mean number of total response opportunities per participant for each treatment	Mean results for each participant	Effect size
Barbetta and Heward (1993)	3 10- to 11-year olds with learning disabilities	Learning geography facts	Number of correct responses on next-day test (out of 7)	Alternating treatments	A: Error-correction B: Error-correction and secondary-response	A: Each response B: Each response	A: 336 B: 496	A: 3.4 words acquired posttreatment B: 5.1 words acquired posttreatment	1.3
Barbetta et al. (1993)	6 8- to 9-year olds with developmental disabilities	Learning to read sight words	Number of correct responses on next-day test (out of 10)	Alternating treatments	A: Error-correction B: Error-correction and secondary-response	A: Each response B: Each response	A: 960 B: 1358	A: 6.1 words acquired posttreatment B: 8.8 words acquired posttreatment	1.3
Drevno et al. (1994)	5 9-year old children	Learning to read science words	Percent of definitions on next-day test	Alternating treatments	A: Error-correction B: Error-correction and secondary-response	A: Each response B: Each response	A: 456 B: 761	A: 29% correct on next-day test B: 41% correct on next-day test	0.6
Johnson, Schuster, and Bell (1996)	5 16- to 17-year olds classified as learning disabled	Learning meanings of science vocabulary	Trials to criterion	Alternating treatments	A: Simultaneous prompting and feedback for correct responses B: Simultaneous prompting and error-correction secondary-response	A: Each response B: Each response	No of response opportunities not reported	A: 49 sessions B: 66 sessions	1.1

Effect sizes. Effect sizes were calculated based on Glass, McGaw and Smith's (1981) formula:

$$\frac{\text{(mean score of secondary-response treatment} - \text{mean score of error-correction treatment)}}{\text{standard deviation of the error-correction treatment}}$$

In all experiments the secondary-response treatment was referred to as the ASR error-correction treatment. Two experiments (Barbetta et al., 1993; Barbetta & Heward, 1993) presented the immediate posttreatment scores for each participant for each treatment and one (Drevno et al., 1994) presented the next-day scores for each participant for each treatment. The posttreatment scores (or the next-day scores) for each participant for a treatment were added and divided by the number of participants to produce a mean posttreatment score (or next-day score) for each treatment. The number of sessions required to reach criterion for each participant in each treatment was added and divided by the number of participants to produce a mean number of trials to criterion for participants in each treatment for Johnson et al. (1996). In this experiment the lower the score the more effective the treatment.

Results and conclusion. All experiments favoured the ASR error-correction procedure. The mean effect size favouring the ASR error-correction treatment was 1.1. The results appear to suggest that providing a secondary-response opportunity following an error is more effective than providing feedback or feedback and an error-contingent model alone. However, in all four experiments the response opportunities were either not controlled or not reported. In the three experiments (Barbetta et al., 1993; Barbetta & Heward, 1993; Drevno et al., 1994) where the number of response opportunities was reported it was found that the error-correction and secondary-response treatment resulted in approximately 45% more practice responses than the error-correction treatment. This raises the possibility that the effectiveness of the error-contingent model is attributable to the additional response opportunities, that is, the additional practice. Barbetta et al. (1993, p. 118) acknowledge this by writing that the error-correction and secondary-response treatment “resulted in students emitting many more responses than in the NR error-correction procedure.” This raises the question as to whether the effectiveness of the secondary-response opportunity would be observed if the number of practice responses was controlled.

Aim. The aim of Experiment 4 was to measure the effects on rate of acquisition and instructional efficiency of providing the participant with an error-contingent model plus a secondary-response requirement with the number of practice responses controlled.

METHOD

Participants and Setting

Twelve Year 2 children participated in the experiments. Table 24 describes the participant characteristics. The participants were screened and identified using the same procedure as was used in Experiments 1, 2 and 3. The experiments were conducted at the same computer stations within the same school as in Experiments 1, 2 and 3.

Learning Tasks

Participants were individually tested on a pool of words selected from the LYC Spelling 6 and/or LYC Spelling 8 programme. Using the same procedure as in Experiment 3, 50 unknown words were identified for each participant.

Pre-Experimental Procedures

The pre-experimental procedures were the same as Experiments 1, 2 and 3. Participants were tested on Spelling 6 words until 50 unknown words were found. However, 50 unknown words were not located for Mark, Jasmine, Miles, Layla, Aimee, Catherine, or Bowen. These participants were tested on the Spelling 8 word list to locate 50 unknown words. They then practised on the typing programme until they could type at 20 correct letters per minute. Participants practised the spelling programme until they could operate it independently. Practice on the typing and spelling programme took two to three sessions.

Table 24

Characteristics of Participants in Experiment 4

Participant	Gender	Age (years, months)	Reading level[a]	Spelling programme employed
1 Mark	Male	7.3	21 (8.0 Years)	Spelling 8
2 Jasmine	Female	6.9	21 (8.0 Years)	Spelling 8
3 Seth	Male	7.2	19 (7.5 Years)	Spelling 6
4 Miles	Male	6.9	19 (7.5 Years)	Spelling 8
5 Layla	Female	6.9	20 (7.5-8.0 Years)	Spelling 8
6 Darcy	Male	6.10	19 (7.5 Years)	Spelling 6
7 Aimee	Female	7.1	20 (7.5-8.0 Years)	Spelling 8
8 Catherine	Female	6.5	23 (8.5 Years)	Spelling 8
9 Johnny	Male	6.10	19 (7.5 Years)	Spelling 6
10 Barry	Male	7.1	15 (6.5 Years)	Spelling 6
11 Dallas	Female	6.4	14 (6.0 Years)	Spelling 6
12 Bowen	Female	6.9	20 (7.5-8.0 Years)	Spelling 8

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

The same letter-pair callibration procedure (White & Haring, 1980) and the same assignment to sets procedure as was used in Experiment 3 was employed to ensure that the two sets of words for each participant were of similar difficulty. As with Experiment 3, the same procedure as was used to construct the two sets of words was also used to replace words correct on the 24-hour probe test within each treatment for each participant. This ensured that the difficulty level of each of the participant's word sets remained equivalent throughout the experiment.

Measurement Procedures

The same measurement procedure as was used in Experiment 3 was used in Experiment 4. Twenty four hours after each practice session (and prior to the next day's practice session) the experimenter individually tested each participant with a 24-hour probe test on the 10 words practised the preceding day. A word correct on the 24-hour probe test was (a) classified as acquired, and (b) removed from the practice set and replaced with a word from the pool of unknown words.

General Procedure

The general procedure was the same as that used in Experiments 1, 2 and 3. Participants were informed at the start of each session which treatment was operating. Except for the last two participants, each treatment lasted five days beginning on a Monday and concluding on a Friday. Practice sessions lasting about 20 minutes began on Monday and concluded on Thursday. Twenty-four hour probe testing began on Tuesday and concluded on Friday.

The experiments involving Dallas and Bowen lasted for up to 9 school days (two school weeks, one of which had a public holiday). In addition, both of these children were absent one day from school during the experiment. The experiment was extended for these participants to see whether the trend that was visible after one week was the same as the trend that was visible after two weeks. Practice sessions for the last two participants began on a Monday and concluded the following Thursday. Twenty-four hour probe testing began on Tuesday and concluded the following Friday.

Aimee practised for two rounds (that is, 20 practice responses instead of 30 practice responses) in both treatments for Sessions 3 and 4 as school finished early due to parent interviews. Apart from the first session, Dallas practised for two rounds in her sessions as she found three rounds too long and became very fatigued.

Computer programming modifications. The computer programmes were adapted so that the computer could measure the number of responses (as opposed to trials) per session. A TRY AGAIN button was created as shown in Figure 34 to allow participants (a) to have a secondary-response opportunity and (b) to have control over the inter-response interval between the primary and secondary response within a trial. A NEXT WORD button was created as shown in Figure 34 to replace the RIGHT ARROW key so that moving between responses was standardized by using buttons.

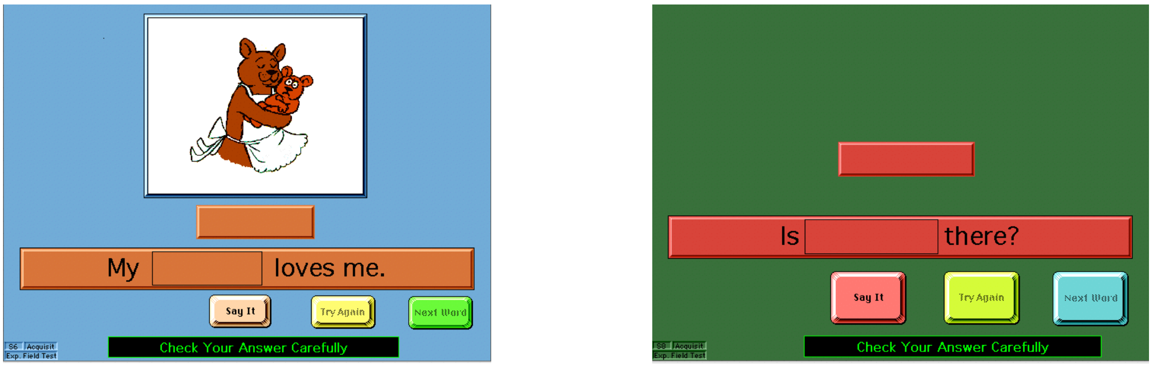


Figure 34. S6 and S8 trial screens with TRY AGAIN and NEXT WORD buttons for the Treatments in Experiment 4.

Error-Contingent Model Treatment. The error-contingent model procedure was the same as that used in Experiment 3. The computer presented (a) the sentence (with no model of the target word) as shown in Figure 34 and (b) the computer-generated voice saying the sentence. If the participant responded correctly the computer provided (a) feedback in the form of a ✔ at the end of the sentence as shown in Figure 35 and (b) the correct answer sound. If the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at the end of the sentence and a model of the word above the sentence as shown in Figure 36 and (b) the incorrect answer sound. Also the NEXT WORD button became highlighted. When ready, the participant clicked the NEXT WORD button to move to the next trial. Except for Aimee and Dallas, the session was completed after 30 trials

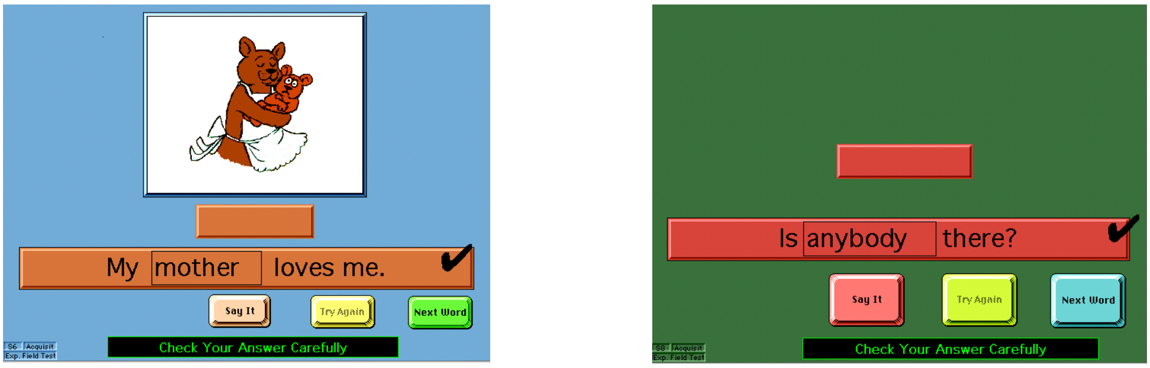


Figure 35. Screen displays for the Spelling 6 and Spelling 8 programmes following a correct response with correct feedback for the Error-Contingent Model Treatment in Experiment 4.

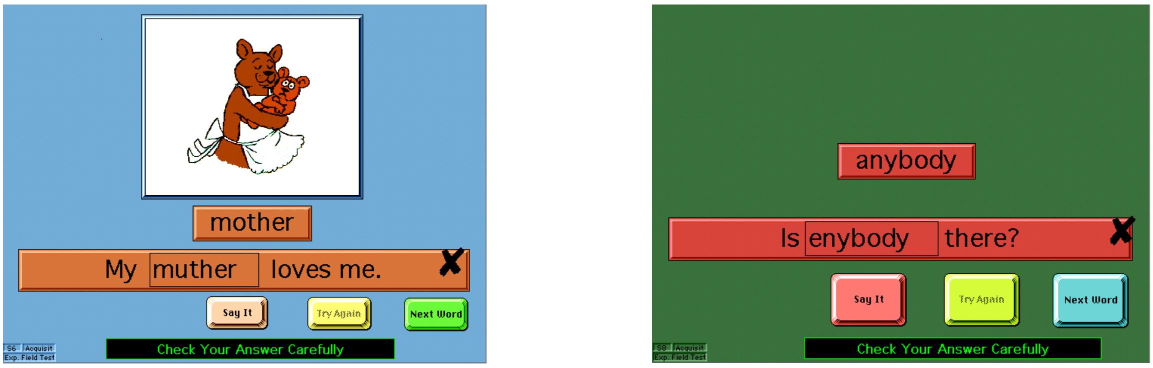


Figure 36. Screen displays for the Spelling 6 and Spelling 8 programmes following an error response with error feedback and an error-contingent model for the Error-Contingent Model Treatment in Experiment 4.

(30 responses) and the experiment was completed after 120 trials (120 responses). Prior to each error-contingent model session the experimenter reminded each participant that the *clue* (model of the target word) would come after they had a go at the word and so it was important to have a go even if they didn't know how to spell the word.

Error-Contingent Model and Secondary Response. Prior to each session participants were reminded that the clue would come after they had a go at a word so it was important to try even if they didn't know the word. The computer presented (a) the sentence as shown in Figure 34 and (b) the computer-generated voice saying the sentence. No model of the target word was presented. If the participant responded correctly the computer (a) provided feedback in the form of a ✔ at the end of the sentence as shown in Figure 35, (b) provided the correct answer sound, and (c) highlighted the NEXT WORD button. However, if the participant responded incorrectly the computer provided (a) feedback in the form of a ✘ at the end of the sentence and a model of the word above the sentence as shown in Figure 37 and (b) the incorrect answer sound. The computer then said the target word and "Try again." The TRY AGAIN button was then highlighted as shown in Figure 37. When the participant clicked on the highlighted TRY AGAIN button the participant's response and the model of the word were removed from the screen. The participant then typed their response (that is, the second response on that trial) and pressed RETURN. Pressing the RETURN key dimmed and deactivated the TRY AGAIN button.

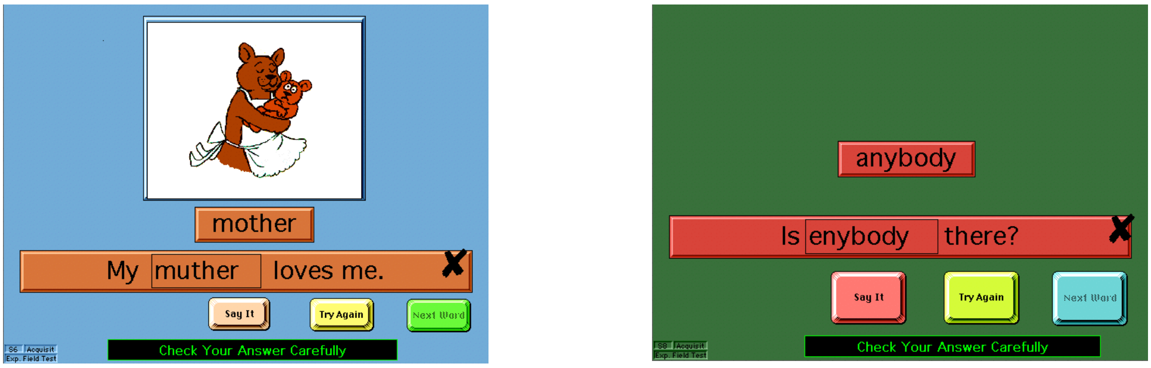


Figure 37. Screen displays for the Spelling 6 and Spelling 8 programmes following an error response (primary response) with error feedback, an error-contingent model and the TRY AGAIN button highlighted for the Error-Contingent Model and Secondary-Response Treatment in Experiment 4.

If the participant responded correctly on the second attempt the computer (a) provided feedback in the form of a ✔ at the end of the sentence as shown in Figure 35, (b) provided the correct answer sound, and (c) highlighted the NEXT WORD button. When ready, the participant clicked this button to move to the next trial. If the participant responded incorrectly a second time the computer (a) provided feedback in the form of a ✘ at the end of the sentence and a model of the word above the sentence as shown in Figure 38, (b) provided the incorrect answer sound, and (c) highlighted the NEXT WORD button.

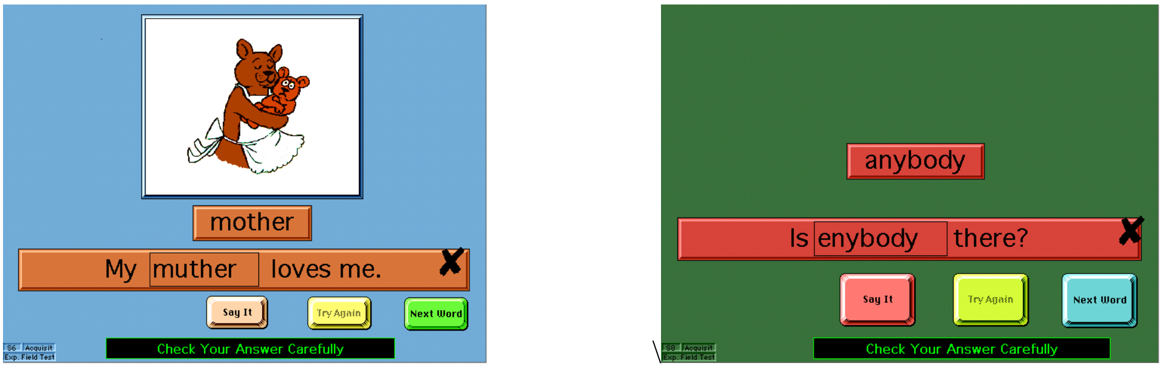


Figure 38. Screen displays for the Spelling 6 and Spelling 8 programmes following an error response (secondary response) with error feedback, an error-contingent model and the NEXT WORD button highlighted for the Error-Contingent Model and Secondary-Response Treatment in Experiment 4.

When ready, the participant clicked this button to move to the next trial. The session was completed once the participant had made 30 responses. As the participant could make up to two responses per trial the number of trials per session varied for each participant depending upon the number of errors made (and therefore the number secondary responses following these errors). The session was completed after 30 responses and the experiment was completed after 120 responses.

Experimental Design

An alternating treatments design (Cooper, Heron, & Heward, 1987) was used. As in Experiment 3, the participants attended two sessions per school day; the Error-Contingent Model Treatment and the Error-Contingent Model and Secondary-Response Treatment. One session was at 9.00 a. m. and the other at 1.30 p. m. Treatment order was counterbalanced so that each participant experienced each treatment twice in the morning and twice in the afternoon. For the last two participants the order of each treatment was arranged so that each treatment was experienced the same number of times in the morning and the afternoon. Experiment 4 was replicated across eleven learners.

RESULTS

Interscorer Agreement

A second-year teacher trainee assessed the accuracy of the scoring of 25% of participants' responses on the probe tests. Interscorer agreement was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying this by 100. Interscorer agreement was 99%.

Procedural Reliability

A second-year teacher trainee conducted a procedural reliability check on 25% of the sessions to ensure that participants received the scheduled experimental treatment. Procedural reliability was calculated by counting the agreements and disagreements between the within-session computer printout and the experimenter's recording sheet of the treatment experienced by each child during each session. The procedural reliability was calculated by dividing the total number of agreements by the total number of agreements and disagreements, and multiplying by 100. Procedural reliability was 100%.

Treatment Integrity

Treatment integrity was assessed by viewing the computer printouts listing every response made, and every prompt received by each participant. These printouts confirmed that each participant was provided with a single response opportunity during every Error-Contingent Model Treatment session and one secondary response opportunity contingent upon an error in every Error-Contingent Model and Secondary-Response Treatment session.

It was concluded that each experimental treatment was implemented as planned for each participant.

Rate of Acquisition

The cumulative number of words correct on the 24-hour probe test for each participant under each treatment is shown in Figure 39. Words correct on the 24-hour probe test were classified as acquired.

Effectiveness

The Error-Contingent Model Treatment was more effective for five participants and the Error-Contingent Model and Secondary-Response was more effective for two participants. There was little difference in effectiveness for four participants one of whom (Barry) acquired almost no words.

Excluding Barry, the number of words acquired after four sessions ranged from 7 to 19 in the Error-Contingent Model treatment, and 5 to 18 in the Error-Contingent Model and Secondary-Response Treatment. This can be seen in Table 25. Barry acquired two words in the Error-Contingent Model Treatment and no words in the Error-Contingent Model and Secondary-Response Treatment. As Barry acquired no words in the Error-Contingent Model and Secondary-Response Treatment, his scores were not included in mean totals in Tables 25 and 26. After six and seven sessions, Dallas and Bowen acquired 14 and 25 words in the Error-Contingent Model Treatment respectively, and 9 and 16 words in the Error-Contingent Model and Secondary-Response Treatment respectively. The mean number of words acquired for all participants after four sessions was 12.2 for the Error-Contingent Model Treatment and 10.4 for the Error-Contingent Model and Secondary-Response Treatment. The difference was not significant ($t = 1.3$, $p < .05$). Mark did not acquire any more words after the second probe test in the Error-Contingent Model and Secondary-Response Treatment while Seth did not acquire any words from the third practice session in either treatment.

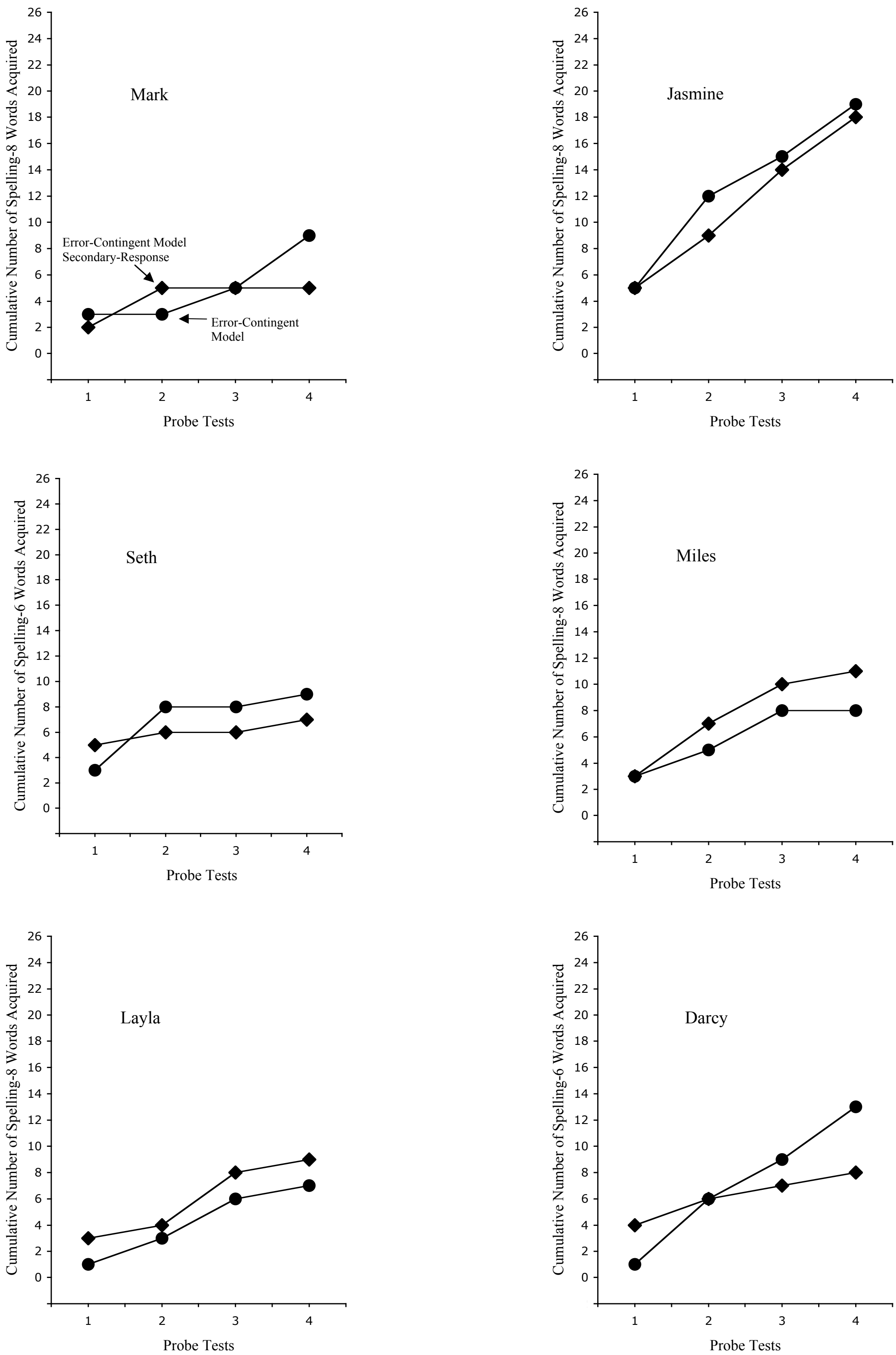


Figure 39. Cumulative number of words correct on the 24-hour probe test for each participant under each treatment.

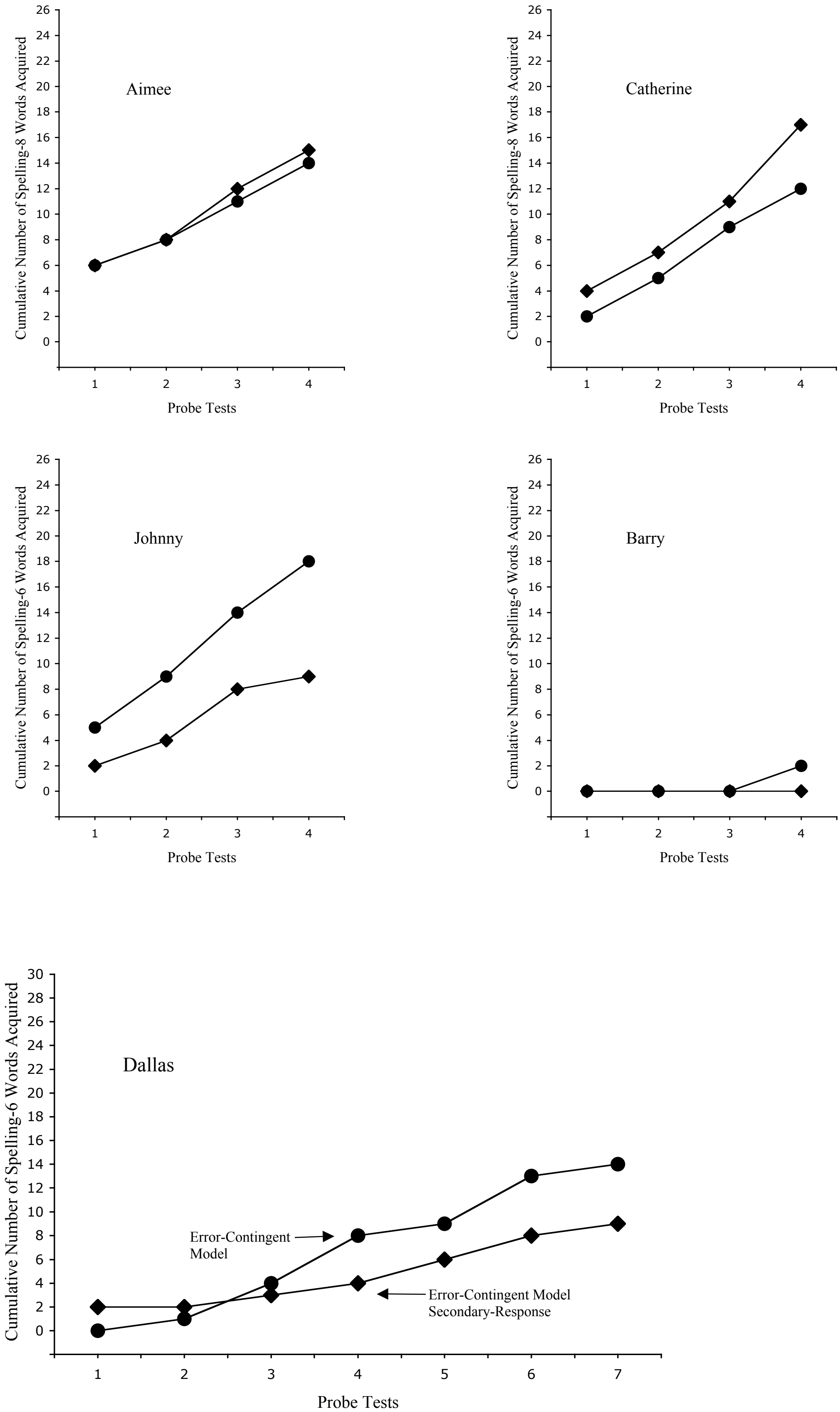


Figure 39. (Continued).

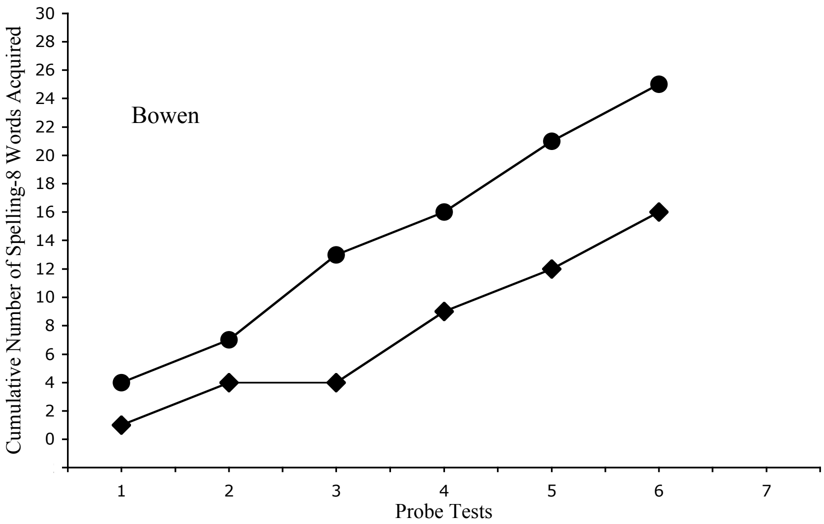


Figure 39. (Continued).

Trials to Criterion

Trials to criterion for each participant was calculated by adding the number of trials required on the acquired words in a treatment and dividing this by the number of words acquired in that treatment. Trials to criterion ranged from 3.1 trials for Aimee to 9.0 trials for Seth in the Error-Contingent Model Treatment, and 1.8 trials for Mark to 4.4 trials for Bowen in the Error-Contingent Model and Secondary-Response Treatment. Across participants, the mean trials to criterion was 5.6 trials for the Error-Contingent Model Treatment and 2.9 trials for the Error-Contingent Model and Secondary-Response Treatment. The difference was significant ($t = 4.9$, $p > .05$).

Responses to Criterion

Responses to criterion for each participant was calculated by adding the number of responses required on the acquired words in a treatment and dividing this by the number of words acquired in that treatment. Responses to criterion ranged from 3.1 responses for Aimee to 9.0 responses for Seth in the Error-Contingent Model Treatment, and 1.7 responses for

Table 25

Effectiveness, Trials to Criterion, Responses to Criterion and Practice Responses to Criterion Data After Four Sessions for Participants Under Error-Contingent Model and Error-Contingent Model and Secondary-Response Treatments

	Effectiveness (number of words acquired after 4 sessions)		Trials to criterion (mean number of trials on acquired words per word acquired)		Responses to criterion (mean number of practice responses on acquired words per word acquired)		Practice responses to criterion (mean number of practice responses on all words per word acquired)	
	Error-Contingent Model	Error-Contingent Model and Secondary -Response	Error-Contingent Model	Error-Contingent Model and Secondary -Response	Error-Contingent Model	Error-Contingent Model and Secondary -Response	Error-Contingent Model	Error-Contingent Model and Secondary-Response
Mark	9	5	7.3	1.8	7.3	1.7	13.3	24.0
Jasmine	19	18	4.9	2.5	4.9	4.0	6.3	6.7
Seth	9	7	9.0	2.7	9.0	6.2	13.4	17.6
Miles	8	11	5.6	3.7	5.6	6.0	15.0	11.1
Layla	7	9	7.1	3.4	7.1	4.8	17.1	13.3
Darcy	13	8	5.7	2.0	5.7	3.5	9.2	15.5
Aimee	14	15	3.1	2.0	3.1	3.1	6.3	6.0
Catherine	12	17	4.6	2.9	4.6	5.2	7.5	6.1
Johnny	18	9	3.8	3.0	3.8	4.7	6.7	13.3
Dallas[a]	9	6	5.0	3.5	5.0	3.7	9.1	14.2
Bowen[a]	16	9	5.6	4.4	5.6	7.8	7.5	13.1
Mean	12.2	10.4	5.6	2.9	5.6	4.6	10.1	12.8
SD	4.2	4.4	1.7	0.8	1.7	1.7	3.9	5.4
Barry	2	0	7.5	-	7.5	-	40.0	-

[a] Data reported after four sessions so that comparisons with other participants could be made.

Mark to 7.8 responses for Bowen in the Error-Contingent Model and Secondary-Response Treatment. Across participants, the mean number of responses to criterion was 5.6 responses for the Error-Contingent Model Treatment and 4.6 responses for the Error-Contingent Model and Secondary-Response Treatments. This difference was not significant ($t = 1.6$, $p < .05$).

Practice Responses to Criterion

Practice responses to criterion was calculated by dividing a participant's total number of practices responses in a treatment by the total number of words that they acquired in that treatment. As Barry did not acquire any words in the Error-Contingent Model Secondary-Response Treatment, instructional efficiency could not calculated for Barry. Practice responses to criterion ranged from 6.3 practice responses per word for Jasmine and Aimee to 17.1 practice responses per word for Layla in the Error-Contingent Model Treatment, and from 6.0 practice responses per word for Aimee to 24.0 practice responses per word for Mark in the Error-Contingent Model and Secondary-Response Treatment. Across participants, the mean practice responses to criterion was 10.1 practice responses per word for the Error-Contingent Model Treatment and 12.8 practice responses per word for the Error-Contingent Model and Secondary-Response Treatment. This difference was not significant ($t = 2.0$, $p < .05$).

Participant's Responses to Instruction

As with Experiment 3, participants were interviewed at the end of the experiment to determine how they responded to instruction. To determine which treatment the participant preferred the experimenter asked, "How did you like to learn your spelling best – moving on the next word or having another go?" To determine how difficult the learner found each treatment the experimenter asked each participant, "When you were working on the computer and went to the next word (or had another go) did you find it easy, middle or hard?" To determine whether the participant enjoyed working on the computer in each treatment the experimenter asked, "When you were working on the computer and you went to the next word (or had another go) did you not enjoy it, enjoy it a bit, or enjoy it a lot?" End-of-experiment social validity data is presented in Table 26.

Table 26

Participants' Social Validity Reports for Experiment 4

	Enjoyment		Difficulty level		Treatment preference
	Error-Contingent Model	Error-Contingent Model and Secondary-Response	Error-Contingent Model	Error-Contingent Model and Secondary-Response	
Mark	A lot	A lot	Easy	Easy	Error-Contingent Model
Jasmine	A lot	A lot	Easy	Easy	Error-Contingent Model
Seth	A lot	A lot	Easy	Easy	Error-Contingent Model and Secondary-Response
Miles	A bit	A lot	Easy	Middle	Error-Contingent Model
Layla	A lot	A lot	Easy	Easy	Error-Contingent Model and Secondary-Response
Darcy	A bit	A bit	Easy	Easy	Error-Contingent Model and Secondary-Response
Aimee	Did not enjoy	Did not enjoy	Hard	Hard	Error-Contingent Model
Catherine	A bit	A bit	Middle	Hard	Error-Contingent Model
Johnny	A lot	A lot	Middle	Hard	Error-Contingent Model and Secondary-Response
Barry	A lot	A lot	Easy	Easy	Error-Contingent Model
Dallas	Did not enjoy	Did not enjoy	Hard	Hard	Error-Contingent Model and Secondary-Response
Bowen	A lot	A lot	Easy	Easy	Error-Contingent Model and Secondary-Response

Mark. Mark acquired nine words in the Error-Contingent Model Treatment, and acquired five words in the Error-Contingent Model and Secondary-Response Treatment, all of which were acquired in the first two sessions. Mark reported that he preferred the Error-Contingent Model Treatment because “You stop worrying about it (the error).” Mark found the presentation of the error feedback and model aversive in both treatments and tended to avoid attending to it. After pressing the RETURN key he watched to see whether he got a ✔or ✘, and then immediately clicked the NEXT WORD button (or, in the Error-Contingent Model Secondary-Response Treatment, the TRY AGAIN button). Because he immediately clicked the NEXT WORD button (or the TRY AGAIN button) the model following an error was in view for less than a second. In the Error-Contingent Model and Secondary-Response Treatment the secondary response was correct on only 36% of occasions. It appears that the secondary response for Mark was a required response (either correct or incorrect) in a response chain that terminated the trial. The reinforcement may therefore not have been contingent on responding correctly but on access to the NEXT WORD button. This button allowed him to escape the trial. However, on the occasions when Mark did attend to the model, the word was often correct on the next trial (or the secondary response) and was then correct 24-hours later. Therefore, the effectiveness of both treatments may have been decreased because Mark found the procedure aversive and did not attend to the prompt.

Jasmine. There was no difference in rate of acquisition between the treatments with Jasmine acquiring 19 words and 18 words in the Error-Contingent Model Treatment and the Error-Contingent Model and Secondary-Response Treatment, respectively. She stated she enjoyed both treatments a lot and found both easy, and she said that she preferred the Error-Contingent Model Treatment because “You get to forget it *(the error)*.”

Seth. Apart from the first probe test, Seth’s rate of acquisition was the same. He acquired nine words in the Error-Contingent Model Treatment and seven words in the Error-Contingent Model and Secondary Response Treatment. He reported enjoying both treatments a lot though stated he preferred the Error-Contingent Model and Secondary-Response

Treatment because "You can get it *(the word)* right" and "I like being right." Despite acquiring fewer words than the mean for all participants, Seth reported that both treatments were easy.

Miles. Miles acquired 8 words in the Error-Contingent Model Treatment and 11 words in the Error-Contingent Model and Secondary-Response Treatment. He reported enjoying the Error-Contingent Model and Secondary-Response Treatment a lot and the Error-Contingent Model Treatment a bit although this was the treatment he reported preferring because if you "Don't get it right, you go onto the next word." The difference in rate of acquisition appears to be due largely to word difficulty during the second session as only two words of the first 10 words in the Error-Contingent Model Treatment were acquired compared to four words of the first 10 words in the Error-Contingent Model and Secondary-Response Treatment during this session. The two additional words acquired in the Error-Contingent Model and Secondary-Response Treatment were smaller words with fewer letter pairs to acquire compared to words in the Error-Contingent Model Treatment.

Layla. Layla acquired seven words and nine words in the Error-Contingent Model Treatment and the Error-Contingent Model and Secondary-Response Treatment, respectively. She stated that she preferred the Error-Contingent Model and Secondary-Response Treatment because "I can think in my head and spell it *(the word)* right" and "I can spell the word."

Darcy. Darcy acquired 13 words in the Error-Contingent Model Treatment and 8 in the Error-Contingent Model and Secondary-Response Treatment. Although acquiring more words under the Error-Contingent Model Treatment, the words that were acquired required a mean of 5.7 responses to acquire whereas the words acquired under the Error-Contingent Model and Secondary-Response Treatment only required a mean of 3.5 responses to acquire. Darcy stated that he preferred the Error-Contingent Model and Secondary-Response Treatment because "I like words right" and "I get to get the words right." An inspection of the words in both treatments showed that word difficulty was well controlled. It is possible that Darcy attended to words differently in each of the treatments. Darcy appeared to use covert self-rehearsal skills during the Error-Contingent Model Treatment but to self-rehearse less often during the Error-Contingent Model and Secondary-Response Treatment.

Aimee. Aimee's rates of acquisition were parallel except for the third day where she learned one more word in the Error-Contingent Model Treatment. She acquired 14 words and 15 words in the Error-Contingent Model and Error-Contingent Model and Secondary-Response Treatments, respectively. Aimee reported that she did not enjoy either treatment. At the beginning of each session she was reminded that participation was voluntary but always wanted to participate.

Catherine. The rate of acquisition was higher for words learned in the Error-Contingent Model and Secondary-Response Treatment (17 words) than in the Error-Contingent Model Treatment (12 words). However, words acquired in the Error-Contingent Model and Secondary-Response Treatment required a greater number of responses to criterion (mean = 5.2) than was the case in the Error-Contingent Model Treatment (mean = 4.6). Catherine reported her preference for the Error-Contingent Model Treatment because you "Don't have to go again" and reported it as "a bit" difficult. This compares to the Error-Contingent Model and Secondary-Response Treatment that Catherine described as "hard." In the classroom the teacher described Catherine as diligent and hard working. It might be that Catherine attends more to tasks that she finds difficult.

Johnny. Johnny learned 18 words under the Error-Contingent Model Treatment with a mean of 3.8 responses to criterion and learned 9 words under the Error-Contingent Model and Secondary-Response Treatment with a mean 4.7 responses to criterion. Johnny reported that he found the Error-Contingent Model and Secondary-Response Treatment "hard" and yet this was the treatment he stated he preferred because he was able to "get to get another try *(after* *an error).*” It is likely that Johnny found the error feedback aversive in both treatments. In the Error-Contingent Model Secondary-Response Treatment he could attend to the model at a level sufficient to respond correctly immediately following the presentation of the error-contingent model. This was not possible in the Error-Contingent Model Treatment as other trials were interspersed between the error response and the next response opportunity. Johnny may therefore have attended more closely to the error-contingent model in this treatment in order to respond correctly on the next trial.

Barry. Barry only acquired two words in the experiment (both in the Error-Contingent Model Treatment). He reported enjoying both treatments a lot and said they were easy. During the Error-Contingent Model Treatment he was often observed attending to the error feedback for less than a second and then pressing the NEXT WORD button. Despite Barry reporting that both treatments were easy, it is clear that the learning task in this experiment was too difficult for him. It was found during the latter days of the experiment that Barry had recently experienced some traumatic events in his home life, and this was likely to have affected his performance during the experiment.

Dallas. Dallas acquired 14 words and 9 words in the Error-Contingent Model Treatment and the Error-Contingent Model and Secondary-Response Treatment, respectively, after seven sessions. Although the Error-Contingent Model Treatment was more effective the slopes were parallel from the fourth session. Dallas reported that she found both treatments “hard” and that she preferred the Error-Contingent Model and Secondary-Response Treatment because “I like getting everything right.” An inspection of the word lists in both treatments found that word difficulty was well controlled. Dallas only practised 20 responses per session from Session 2 onwards because she found the first session with 30 practice responses too long. It might be the slope of the graph in the Error-Contingent Model Treatment decreased from Session 4 because the spelling session was becoming aversive for

Dallas. As a result she decreased her attending to the error-contingent models in both treatments and increased her attending to other children and to other non-experimental stimuli during each session as the experiment continued. Prior to each session Dallas was reminded that participation was voluntary and she always stated that she wanted to do the spelling session. However, it seemed that Dallas wanted to participate in each session because it was reinforcing to leave the classroom to participate in the experiment, engage one-on-one with the experimenter and play a computer game.

Bowen. Bowen's rate of acquisition was 25 words in the Error-Contingent Model Treatment and 16 words in the Error-Contingent Model and Secondary-Response Treatment after six sessions. Bowen was often observed to talk to herself during sessions. During both treatments she was often observed to say, "I'm really trying hard to get this right" when she was responding, and was often observed to say, "Yay" and "I got a tick" when she responded correctly. The data paths were parallel except for the third probe test where no words were acquired under the Error-Contingent Model and Secondary-Response Treatment. The experimenter visually inspected the words for each treatment for the third session and concluded that there were three words that Bowen might have acquired that session. All three of these words were acquired the following session so it is likely that word difficulty accounted for the difference in rate of acquisition for this session. Bowen reported that she found both treatments "easy" and enjoyed them "a lot," and she selected the Error-Contingent Model and Secondary-Response Treatment when asked which one she preferred most.

Effects of the Models

The percentage of responses correct following the presentation of a model (either in the error-contingent model position or the antecedent position for a secondary response) in the first round of all four sessions (the first presentation of a word in a session) for the Error-

Contingent Model and Error-Contingent Model and Secondary Response Treatments can be seen in Table 27. The percentage of responses correct following the error-contingent model

Table 27

Percentage of Words Correct Following the Error-Contingent Model in (a) the Error-Contingent Model and (b) the Error-Contingent Model and Secondary-Response Treatments

	Percentage of words correct following the first presentation of the model					
Name	Error-Contingent Model			Error-Contingent Model and Secondary-Response		
	Acquired words	Non-acquired words	All words	Acquired words	Non-acquired words	All words
Mark	25	0	13	25	32	31
Jasmine	47	0	33	74	30	61
Seth	17	4	13	100	31	47
Miles	42	37	39	80	41	28
Layla	7	4	5	80	57	67
Darcy	50	18	38	29	64	55
Aimee	60	32	44	93	64	81
Catherine	47	33	41	70	29	62
Johnny	60	13	37	77	33	52
Dallas	27	32	29	93	38	57
Bowen	31	0	28	76	67	73
Mean	38	16	29	76	44	56
Barry	25	7	9	0	39	39

for the first round of a session was calculated using the same procedure as in Chapter 7. That is, the total number of error-contingent models presented on the first round of a session was counted. The number of responses correct following the error-contingent model was calculated by counting the number of words correct on the second round of a session which had been incorrect (and had received an error-contingent model) on the first round of a session. This was converted to the percentage of responses correct following the presentation of an error-contingent model during the first round of a session. The percentage of responses correct following the error-contingent model for the secondary response was calculated by counting the number of secondary responses correct following an error-contingent model on the first round of a session. This was then converted to a percentage.

In the Error-Contingent Model Treatment, the effectiveness of the error-contingent model in the first round ranged from 5% for Layla to 44% for Aimee. The mean percentage of correct secondary responses following the error-contingent model for the first round of a session was 29%. It can be seen that the mean percentage of correct secondary responses following the model was higher for acquired words (38%) than non-acquired words (16%).

In the Error-Contingent Model and Secondary-Response Model Treatment, the effect of the error-contingent model in the first round ranged from 28% correct secondary responses for Miles to 81% for Aimee. The mean percentage of correct responses following the error-contingent model in this treatment for the first round of a session was 56%. The mean percentage of correct responses following the error-contingent model for the secondary response was higher for acquired words (76%) than non-acquired words (44%).

DISCUSSION

Both the Error-Contingent Model and Error-Contingent Model and Secondary-Response Treatments were effective in teaching young children how to spell new words on the computer. The experiment showed that participants acquired an average of 12.2 words in the Error-Contingent Model Treatment and 10.4 words in the Error-Contingent Model and Secondary-Response Treatment as a result of 120 practice responses.

Effectiveness

The Error-Contingent Model Treatment was slightly more effective than the Error-Contingent Model and Secondary-Response Treatment. This was a surprising result because it was thought prior to the experiment that this treatment would be less effective on the grounds that previous experiments favoured the secondary-response treatment (Barbetta et al., 1993; Barbetta & Heward, 1993; Drevno et al., 1994; Johnson et al., 1996). The present experiment differs from those of Barbetta et al. (1993), Barbetta and Heward (1993), Drevno et al. (1994) and Johnson et al. (1996) in that it controlled the number of practice responses, and not the number of trials. On average, the participants in these four experiments received 14 percent (Barbetta et al., 1993) to 67 percent (Drevno et al., 1994) more response opportunities during the secondary-response treatments than they did in the error-correction only treatments. The present experiment demonstrates that it is not the secondary responses which produce the superior effect but the additional practice because when the number of practice responses is controlled as it was in the present experiment the secondary response treatment results in no increase in rate of acquisition whatsoever.

Trials to Criterion

As expected, the Error-Contingent Model and Secondary-Response Treatment required fewer trials to criterion on average than the Error-Contingent Model Treatment. It was not possible to control both the number of trials to criterion and the number of responses to criterion as the number of responses per trial differed from trial to trial in the Error-Contingent Model and Secondary-Response Treatment. This result then is a function of the lack of experimental control over the number of trials to criterion. Because the number of trials is no longer controlled, trials to criterion is no longer a meaningful measure of instructional efficiency.

Responses to Criterion

Although almost equally effective, the Error-Contingent Model and Secondary-Response Treatment required fewer responses per word acquired for 6 of the 11 children and slightly fewer responses to criterion on average. Although the presentation of a secondary-response opportunity may have been aversive for some participants, it appears that when participants did attend to the model the attending was higher than in the Error-Contingent Model Treatment.

Practice Responses to Criterion

The Error-Contingent Model Treatment required fewer total practice responses per word acquired in 6 out of 11 cases. When the number of practice responses is controlled, the treatment which is more effective also turns out to be the more efficient although the differences were slight.

Error Corrections as Aversive Stimuli

Participant's reports and comments on treatment preferences suggest that all participants found error corrections aversive. Mark, Jasmine, Miles, Aimee and Barry often managed the error correction in the Error-Contingent Model Treatment by moving to the next trial to escape the error feedback. This is probably what Mark meant when he said that, "You stop worrying about it (the error)" and Jasmine meant when she said, "You get to forget it (the error)." Catherine, Johnny and Bowen also found the error correction aversive, but rather than escaping the trial, they attended to the model to avoid the error on the next round. This is probably what Bowen meant when she said, "I am trying really hard to get this right." Seth, Layla, Darcy and Dallas preferred the Error-Contingent Model and Secondary-Response Treatment possibly because responding correctly on the secondary response often generated positive feedback. This is probably what Seth meant when he said, "You can get it (the word) right" and "I like being right." Layla also reported that, "I can think in my head and spell it (the word) right" and "I can spell the word" while Darcy reported, "I like words right."

Summary

The results of Experiment 4 differ from those of previous experiments. The experiment found that the Error-Contingent Model Treatment was just as effective for six participants and more effective for five participants although it did require a greater number of practice responses per word acquired than the Secondary-Response Treatment and was, therefore, less efficient.

REFERENCES

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition and maintenance of geography facts by elementary students with learning disabilities. *Journal of Behavioral Education, 3*, 217-233.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus: Merrill Publishing Company.

Drevno, G. E., Kimball, J. W., Possi, M. K., Heward, W. L., Gardner, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and generalization of science vocabulary by elementary students: A systematic replication. *Journal of Applied Behavior Analysis, 27*, 179-180.

Espin, C. A., & Deno, S. L. (1989). The effects of modeling and prompting feedback strategies on sight word reading of students labeled learning disabled. *Education and Treatment of Children, 12*, 219-231.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.

Heward, W. L. (1994). Three "low tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 283-320). California: Brooks/Cole Publishing Company.

Johnson, P., Schuster, J., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6*, 437-458.

Kulhavy, R., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522-528.

Nelly, E., & Smith, A. (2000). *PM Benchmark kit: An assessment resource for emergent - 12 years reading age*. Melbourne: Nelson ITP.

Singh, N. N., Winton, A. S., & Singh, J. (1985). Effects of delayed versus immediate attention to oral reading errors on the reading proficiency of mentally retarded children. *Applied Research in Mental Retardation, 6*, 283-292.

Skinner, C. H., Shapiro, E. S., Turco, T. L., Cole, C. L., & Brown, D. K. (1992). A comparison of self- and peer-delivered immediate corrective feedback on multiplication performance. *The Journal of School Psychology, 30*, 101-116.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching (2nd ed.)*. Ohio: Charles E. Merrill Publishing Company.

CHAPTER 9

SUPPLEMENTARY ANALYSIS OF THE RESULTS OF THE FOUR EXPERIMENTS

Four experiments have been reported in this thesis. Experiment 1 attempted to examine the relationship between accuracy level during instruction and rate of acquisition. This experiment failed because prompting was poorly controlled. Experiment 2 was a repetition of Experiment 1 with better control over the number of prompts employed. Experiment 3 compared the effects of an antecedent model used as a prompt against the effects of an error-contingent model used as an error correction. Experiment 4 compared the effects of an error-contingent model against the effects of an error-contingent model plus a secondary-response opportunity. These experiments all employed highly similar, computer controlled, experimental procedures and produced patterns of acquisition which shared a number of common features. The aim of this supplementary analysis is to review and compare the data on rate of acquisition, responses to criterion, practice responses to criterion, and individual variability across the four experiments.

In order to compare data across the four experiments, several procedural differences needed to be taken into account. First, the experimental treatments in Experiment 1 were up to 10 sessions in length, the treatments in Experiment 2 were up to 7 sessions in length, and the treatments in Experiments 3 and 4 were, with two exceptions, 4 sessions in length. In order to make comparisons that were based on the same number of sessions and almost the same number of responses per participant per treatment across the four experiments, the analysis which follows makes use of just the data from the first four days of Experiments 1 and 2.

Second, Experiments 1 and 2 measured acquisition using 24-, 48- and 72-hour postinstructional probe tests whereas Experiments 3 and 4 measured acquisition using a single 24-hour daily probe test. In order to compare rates of acquisition, acquisition was recalculated in Experiments 1 and 2 using just the data from the 24-hour probe test.

Third, the mean accuracy levels during instruction for treatments in Experiments 1 and 2 (reported in Chapter 5) were originally calculated on all practice sessions in these experiments. As Experiments 1 and 2 lasted longer than four sessions, the mean accuracy level during instruction was recalculated for just the first fours days of practice. The mean accuracy levels during instruction after four sessions for each treatment in Experiments 1 and 2 are shown in Table 28.

Table 28.

Mean Accuracy Level During Instruction of Each Treatment in Experiments 1 and 2 After Four Sessions

Experiment	Treatment	Mean accuracy level
1	65% target accuracy level	63
	80% target accuracy level	62
	95% target accuracy level	66
Expt mean		64
SD		2
2	65% target accuracy level	69
	80% target accuracy level	76
	95% target accuracy level	79
Expt mean		75
SD		5

Fourth, in order to compare the mean number of responses per word acquired across treatments within experiments the mean number of responses to criterion was calculated for Experiment 3 and for the first four days of Experiments 1 and 2.

Fifth, the number of correct responses to criterion in Experiments 1, 2, 3 and 4 for the first four days was calculated so that the number of correct responses to criterion could be analysed. The mean number of correct responses to criterion in each treatment was measured by counting the number of correct responses to criterion for each participant in each treatment and dividing this by the number of words acquired by that participant. The mean number of correct responses to criterion for each participant for each treatment was summed and divided by the number of participants in that treatment. This produced a mean number of correct responses to criterion for the first four days for each treatment.

Sixth, in order to compare the mean effectiveness (mean number of words acquired) across each treatment within the four experiments, mean effectiveness was recalculated for Days 1 to 4 of Experiments 1 and 2.

Finally, instructional efficiency (that is, the mean number of practice responses on all words per word acquired) was calculated for Days 1 to 4 of Experiment 1. Instructional efficiency could not be calculated for Experiment 2 because only the first 10 of the words practised were tested on the 24-hour probe test.

PART ONE: THE DEGREE OF STIMULUS CONTROL EXERCISED BY THE MODEL

It was decided to compare the effectiveness of the model of the correct spelling across experiments (a) in order to identify any commonalities between treatments where the model was in the same position within a trial, and (b) because comparisons could be made between treatments in different experiments that could not be made between treatments within the same experiment.

Experiment 2 was not included in the analysis because there was a ceiling effect on the words acquired. In order to make comparisons across Experiments 1, 3 and 4 it was necessary to recalculate the effects of the model on the first round responses in each treatment in Experiment 1 for the first four sessions. Table 29 shows the mean percentage of spelling responses correct following the presentations of the model for the first round in each session in each treatment after four sessions. For ease of comparison, antecedent-model treatments are listed first, followed by error-contingent model treatments. Treatments that provided a model immediately prior to responding generated a mean of 61% correct responses for all words following the model compared to a mean 26% of responses correct for all words following the

Table 29.

Comparison of Mean Percentage of Words Correct Following the Model for the First Round

Expt	Treatment	Mean percentage of words correct following the first presentation of the model in a session		
		Acquired words	Non-acquired words	All words
1[a]	65% Mean Target Accuracy Level	66	52	64
1	80% Mean Target Accuracy Level	66	55	61
1	95% Mean Target Accuracy Level	58	46	53
3	Antecedent Model	81	56	70
4	Error-Contingent Model and Secondary-Response (antecedent model for secondary response)	72	44	56
	Mean	69	51	61
	SD	9	5	7
3	Error-Contingent Model	26	9	23
4[b]	Error-Contingent Model	38	16	29
	Mean	32	13	26
	SD	8	5	4

[a] The spelling task was too difficult for Lee and he was subsequently removed from the experiment after one treatment. His data is not included in the table.
[b] Barry did not acquire any words in the Error-Contingent Model and Secondary-Response Treatment. His data is not included in the table.

model in the two Error-Contingent Model Treatments in Experiments 3 and 4. Within each treatment it can be seen that the percentage of words correct was higher for acquired words than non-acquired words. The mean percentage of words correct following the model was higher in the Antecedent-Model Treatment in Experiment 3 than either of the antecedent-model treatments in Experiment 1 or the antecedent model for the secondary response in the Error-Contingent Model and Secondary-Response Treatment in Experiment 4.

Discussion

In every treatment the level of stimulus control exercised by the model during practice was greater for acquired words than non-acquired words. This was not surprising given that a correct response following a model was an opportunity for transfer of stimulus control.

The mean percentage of words correct following the presentation of a model of the correct response was higher in the Antecedent-Model Treatment in Experiment 3 (70%) than in any of the treatments in Experiment 1 (mean 59%). The lower level of correct responding in Experiment 1 may have been because the children quickly discovered that an error-contingent model was also presented within a trial. It will be recalled from Chapter 5 that this probably affected the attending of participants with some children making careless errors in the knowledge that another opportunity to view the correct spelling would occur following an error.

It was expected that the error-contingent model in Experiment 4 would prompt a correct secondary response as frequently as the antecedent model prompted correct responding in the Antecedent-Model Treatment in Experiment 3, that is, 70% of the time. However, the error-contingent model only prompted a correct secondary response on 56% of occasions even though the latency between the model and the response was minimal in both cases. One possible reason for this is the fact that some of the children found the error feedback aversive. Some of the children were observed to attend less to the model in the error-contingent position in Experiment 4 than was the case with the antecedent model in Experiment 3. Interestingly, participants in the Error-Contingent Model Secondary-Response Treatment required fewer practice responses (4.6 on average) to acquire a new spelling response than was the case for participants in the Antecedent-Model Treatment in Experiment 3 (5.5 on average). This effect remains unexplained and warrants further experimental analysis. One possibility is that the model in the Error-Contingent Model Secondary-Response Treatment allows participants to focus their attention on the error component of the word because, in this position, they are able to compare the letter pairs in their incorrect response to the letter pairs in the postcedent model – something which they cannot do with an antecedent model.

The mean percent of words correct following the antecedent model was 61% whereas the mean percent of words correct following the error-contingent model was 26%. Following the results of Experiment 3, the higher mean percent of words correct following the model presented immediately before the practice response rather than immediately following an error response was expected. This was because the antecedent model gained stimulus control over the participant's spelling response more often during instruction than did the model in the consequent position because of the simultaneous presentation of the practice stimulus and the presentation of the antecedent model. In the error-contingent model treatments the latency between the presentation of the error-contingent model and the presentation of the practice stimulus was anywhere from approximately 30 seconds to 14 minutes.

PART TWO: OVERALL EFFECTIVENESS AND EFFICIENCY

Table 30 shows the mean effectiveness, mean practice responses to criterion, mean trials to criterion, mean responses to criterion and mean correct responses to criterion across all treatments in all experiments after four sessions.

The mean effectiveness (number of words acquired) after four sessions was 13.1 words for Experiment 1, 9.9 words for Experiment 2, 11.5 words for Experiment 3 and 11.3 words for Experiment 4. Mean effectiveness over the four experiments after four sessions was 12.1 words (SD = 1.4) acquired over four 30-response practice sessions.

The mean number of practice responses on all words practised per word acquired after four sessions was 9.4 practice responses for Experiment 1, 10.0 practice responses for Experiment 3, and 11.5 practice responses for Experiment 4. Due to the methodology employed in Experiment 2 effectiveness had a ceiling of 10.0 responses and practice responses to criterion could not be calculated. The mean practice responses to criterion over Experiments 1, 3 and 4 was 10.1 practice responses (SD = 1.5).

Table 30

Mean Effectiveness, Mean Practice Responses to Criterion, Mean Trials to Criterion, Mean Responses to Criterion and Mean Correct Responses to Criterion Across All Treatments in Experiments 1, 2, 3 and 4 After Four Sessions

Expt	Treatment	Mean effectiveness (mean number of words acquired)	Mean practice responses to criterion (mean number of practice responses on all words per word acquired)	Mean trials to criterion (mean number of trials on acquired words per word acquired)	Mean responses to criterion (mean number of practice responses on acquired words per word acquired)	Mean correct responses to criterion (mean number of correct practice responses on acquired words per word acquired)
1[a]	63% accuracy level	14.8	7.8	4.9	4.9	3.3
	62% accuracy level	12.3	10.1	4.7	4.7	3.1
	66% accuracy level	12.1	10.2	5.2	5.2	3.5
	Expt mean	13.1	9.4	4.9	4.9	3.3
	SD	1.5	1.4	0.3	0.3	0.2
2	69% accuracy level	10.0 [b]	-[c]	4.5	4.5	3.6
	76% accuracy level	9.9	-	4.5	4.5	3.4
	79% accuracy level	9.8	-	4.4	4.4	3.3
	Expt mean	9.9	-	4.5	4.5	3.4
	SD	0.1		0.1	0.1	0.2
3 [d]	Antecedent Model	12.0	9.4	5.5	5.5	4.2
	Error-Contingent Model	10.9	10.6	5.4	5.4	1.9
	Expt mean	11.5	10.0	5.5	5.5	3.1
	SD	0.8	0.8	0.1	0.1	1.6
4[e]	Error-Contingent Model	12.2	10.1	5.6	5.6	2.5
	Error-Contingent Model and Secondary-Response	10.4	12.8	2.9	4.6	3.3
	Expt mean	11.3	11.5	4.3	5.1	2.9
	SD	1.3	1.9	1.9	0.7	0.6
Mean		12.1	10.1	4.8	4.9	3.2
SD		1.4	1.5	0.8	0.5	0.6

[a] The spelling task was too difficult for Lee and he was subsequently removed from the experiment after one treatment. His data was not included in the table.

[b] Because only the first 10 words were measured for acquisition, a ceiling effect occurred. Therefore the mean effectiveness of Experiment 2 was not used to calculate total mean effectiveness of all experiments.

[c] As only the first 10 words were measured for acquisition, a ceiling effect occurred. Instructional efficiency could therefore not be calculated.

[d] As Jeff's data was incomplete in Experiment 4, his data was not used to calculate mean effectiveness.

[e]Barry did not acquire any words in the Error-Contingent Model and Secondary-Response Treatment. His data was not included in the table.

The mean trials to criterion (mean number of trials on acquired words per word acquired) after four sessions was 4.9 trials for Experiment 1, 4.5 trials for Experiment 2, 5.5 trials for Experiment 3 and 4.3 trials for Experiment 4. The mean trials to criterion over all experiments was 4.8 trials (SD = 0.8).

As there was only one response per trial in Experiments 1 to 3, the mean number of responses to criterion was the same as the mean number of trials to criterion. The mean responses to criterion for Experiment 4 was 5.1 responses. The mean number of responses to criterion over all experiments was 4.9 responses (SD = 0.5). Across all experiments 70% of participants required between four and six practice responses per word acquired and 95% of participants required between three and seven practice responses per word acquired. The variability between participants was therefore small.

The mean number of correct responses per word acquired for treatments that contained an antecedent model (all treatments in Experiments 1 and 2, the Antecedent-Model Treatment in Experiment 3, and the Error-Contingent Model and Secondary-Response Treatment in Experiment 4) was 3.5 correct responses (SD = 0.3). For the Error-Contingent Model Treatments in Experiments 3 and 4, the mean number of correct responses to criterion was 2.2 correct responses (SD = 0.4). The variability across learners on this measure was extremely small.

Discussion

Practice responses to criterion. Overall, participants required about 10 total practice responses to acquire a new spelling response. The antecedent-model treatments in Experiments 1 and 3 were slightly more efficient than the error-contingent model treatments in Experiments 3 and 4. It will be recalled from Chapter 5 that the findings of Experiment 2 suggested that teaching procedures that produced a higher proportion of correct responses during practice also produced a higher rate of acquisition.

The error-contingent model treatments in Experiments 3 and 4 were more efficient than the Error-Contingent Secondary-Response Treatment in Experiment 4. This was unexpected because previous experiments favoured the secondary-response treatment (Barbetta, Heron, & Heward, 1993; Barbetta & Heward, 1993; Drevno et al., 1994; Johnson, Schuster, & Bell, 1996). These experiments, however, did not control the number of practice responses. The present experiment demonstrates that it is not the secondary responses which produces the superior effect but the additional practice because when the number of practice responses is controlled as it was in the present experiment the secondary-response treatment results in no increase in rate of acquisition.

Trials to criterion and responses to criterion. The mean number of trials to criterion across all experiments, that is, the mean number of trials required in order to acquire a response, was just over four trials. There was almost no variability across treatments except for the Error-Contingent Model and Secondary-Response Treatment in Experiment 4 which was the only treatment where the number of trials differed from the number of practice responses. Across all treatments in all experiments, participants, on average, required about five responses to criterion, that is, five practice responses in order to acquire a response. Across all treatments there was very little variability in this finding. This suggests that it may not be the number of practice trials that is important for acquisition but rather the number of practice responses. This will be discussed further in Chapter 10.

PART THREE: INDIVIDUAL VARIABILITY

Table 31 shows the characteristics of all participants in all treatments across the experiments. Participants ranged in age from 6.0 years to 7.6, and reading ages ranged from 6.0 years to 10.0 years. The total number of participants with complete data in the four experiments was 40.

Table 31.

Characteristics of Participants in Each of the Four Experiments

Expt	Number of participants	Age range of participants (years, months)	Reading level range of participants[a]
1	9	6.0 - 6.9	6 –18 (6.0 –7.5 Years)
		(Mean = 6.5)	Mean = 11 (6.0 Years)
2	9	7.1 – 7.6	21-26 (8.0-10.0 Years)
		(Mean = 7.3)	Mean = 23 (8.5 Years)
3	11	6.3 – 6.10	15-23 (6.5-8.5 Years)
		(Mean = 6.6)	Mean = 20 (8.0 Years)
4	11	6.4 – 7.3	14-23 (6.0-8.5 Years)
		(Mean = 7.0)	Mean = 19 (7.5 Years)

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

The means presented in Table 30 hide the fact that there were differences in the way in which individual children responded to instruction.

Experiment 1. Lee only participated in one treatment because he required more than twice as many practice responses to learn a new spelling response than the mean of 5.6 practice responses required by the other participants in Experiment 1. This was almost certainly due to the fact that the learning task for Lee was more difficult than for the other children in Experiment 1. As stated in Chapter 5, 60% of the letter pairs in Lee's spelling words were unknown prior to the experiment compared with only 40% of unknown letter pairs in words studied by the other children in Experiment 1.

Experiment 2. Two participants (Carlos and Tania) acquired 10 spelling responses with mean of 3.6 responses to criterion compared to the mean of 4.5 responses to criterion required by the Experiment 2 children in general. This faster acquisition rate is likely due to the attending and self-rehearsal skills of these children. Informal observations of these participants showed that they interacted little with other participants, and when they did it was at the end of a trial rather than during it. That is, their attending during the trial was high. Both were observed to attend closely to the model of the correct spelling and Carlos both verbally self-rehearsed the spelling several times and often identified the particular letter-pair errors which he had made when he misspelled a word. In addition, both of these participants had reading ages about 2.5 years above their chronological age which means that their knowledge of phoneme-grapheme relationships was similarly advanced.

Experiment 3. Darryl, Joel and Catherine each acquired about 19 words for the Antecedent-Model Treatment and about 17 words in the Error-Contingent Model Treatment compared to the mean of about 12 words and 11 words, respectively, acquired by the 11 children as a group. These participants had reading ages about 1.5 years above their chronological age and presumably had acquired a similarly advanced level of sound-spelling knowledge.

Experiment 4. Barry acquired only two words in the Error-Contingent Model Treatment and no words in the Error-Contingent Model and Secondary-Response Treatment. It will be recalled from Chapter 8 that Barry had recently experienced some traumatic events in his home life. It also appeared that the spelling task was too difficult for him. Jasmine acquired 19 words in the Error-Contingent Model Treatment and 18 in the Error-Contingent Model and Secondary-Response Treatment. As with the high-achieving participants in Experiment 2, informal observations suggest that Jasmine exhibited high levels of attention to the model and feedback and entered the experiment with above average levels of sound-spelling knowledge.

Error-Contingent Events as Aversive Stimuli

As the experiments proceeded it became apparent that many of the participants found the error feedback aversive. In fact, the majority of the children did not like error consequences. Error consequences were so aversive for one child (Alice, Chapter 5) that she was withdrawn from Experiment 2 because several times she refused to attempt to spell the target word in case she got it wrong, and then started crying when she did respond incorrectly. Other children complained to themselves when they responded incorrectly or, in Experiment 2, pressed the right arrow key immediately following the presentation of the error feedback to escape to the next trial. Escaping the trial also seemed to be a strategy that a number of children used in the error-contingent model treatments in Experiments 3 and 4. This they did by immediately clicking the NEXT WORD button instead of attending to the correct spelling when presented with the error-contingent model. Other children appeared to have been motivated to avoid future errors by attending closely to the correct spelling when presented with the error-contingent models in these experiments.

Differences Between Highest Achieving and Lowest Achieving Participants

The final grouped analysis undertaken in this section involved a comparison between the highest achieving and the lowest achieving children across the three usable experiments. The number of words a participant acquired under each treatment within an experiment (Experiments 1, 3 and 4) after four sessions was identified. The two participants whose acquisition scores were highest within each experiment were selected as the highest achieving participants and the two participants with complete acquisition data whose scores were the lowest within each experiment were selected as the two lowest achieving participants. Characteristic of these 12 children are shown in Table 32.

It can be seen that the mean age of participants in the high-performance group was two months younger than the mean age of participants in the low-performance group. It can also be seen that the mean reading age of the high-performance group was 14 months higher than that of the low-performance group. Interestingly, the mean reading age of the low-

Table 32.

Comparisons of Characteristics of Six High Performance and Six Low Performance Participants from Experiments 1, 3 and 4

Expt	Name	Age (in months)	Relationship to chronological reading age[a] (in months)	Mean effectiveness (mean number of words acquired after 4 sessions per expt)	Mean responses to criterion (mean number of practice responses on acquired words after 4 sessions per expt)
High achieving					
1	Sade	80	7 above	14.0	5.0
	Gema	72	6 above	14.6	4.2
3	Joel	80	19 above	21.5	4.6
	Catherine	75	24 above	18.5	4.5
4	Jasmine	81	15 above	18.5	4.5
	Catherine	77	22 above	14.5	4.9
Mean		78	16 above	16.9	4.6
Low achieving					
1	Mark	78	1 above	9.5	6.5
	Scarlett	79	2 below	10.0	5.4
3	Leigh	78	6 above	6.5	5.4
	Leon	81	3 above	7.0	6.2
4	Mark	87	9 above	7.0	4.5
	Dallas	76	4 below	7.5	4.4
Mean		80	2 above	7.9	5.4

[a]Benchmark Reading Kit (Nelly & Smith, 2000)

performance group was approximately the same as the average for their age. The high-performance group, on average, acquired about 17 words after four sessions while participants in the low-performance group, on average, acquired only half as many words over the same time. The difference in the number of responses required by each group was small. The high-performance group required an average of 4.6 practice responses per words acquired while participants in the low-performance group required an average of 5.4 practice responses per word acquired.

Participants in the high-performance group acquired over twice as many words as participants in the low-performance group. Several reasons for this difference are explored. In order to determine whether this difference was due to participants in the high-performance group having fewer letter pairs to acquire than participants in the low-performance group, the numbers of letter-pair errors on the first 10 words for each participant in each treatment in Experiments 3 and 4 were counted. Experiments 3 and 4 were selected for this analysis because each treatment was matched for letter-pair errors for each participant. The results of this analysis revealed that participants in the high-performance groups in Experiments 3 and 4 had a mean of 30 letter pairs to acquire while participants in the low-performance groups had a mean of only 25 letter pairs to acquire. The difference in performance between the two groups was therefore not due to task difficulty, that is, the number of letter pairs to be acquired. The mean age of participants in the high-performance group was two months younger than the mean age of participants in the low-performance group so the difference in performance was not due to participants being older. The difference in performance during the experiments may be due to individual learning history effects. It seems likely that participants in the high-performance group were more knowledgeable with respect to their knowledge of phoneme-grapheme relationships. For example, all of the participants in the high-performance group knew all the main letter-sound (grapheme-phoneme) relationships and appeared fluent when sounding out letters. Since many of these relationships are reversible, it is likely that these children heard the letter sound (phoneme) and then typed the corresponding letter (grapheme). This was probably not the case with participants in the low-performance group. Leon, in the low-performance group, self-rehearsed on one occasion and stated, "What sound does that *(letter)* make again?" So, while participants in the high-performance group each had no fewer letter pairs to acquire, it seems that the task difficulty was lower due to their higher level of alphabet knowledge. This seems to be an example of the Matthew Effect (Stanovich, 1986) where more able students start with more background knowledge and end up learning more than the less able students (Nuthall, 2001).

In addition to having higher levels of alphabet knowledge the high-performance group also appeared to have better developed learning-to-learn skills. As a group, children in the high-performance group attended to the model of the correct spelling more carefully and more consistently than children in the low-performance group. The high-performance participants self-rehearsed the correct spelling and, when in the consequent position, compared the error component of their response against the correct spelling in the error-contingent model rather than just looking at the model as participants in the low-performance tended to do. Again, these results are consistent with Nuthall's (2001) argument that more able students are more persistent, less likely to be distracted, and more likely to ask questions. He suggested that a key difference between high-performance and low-performance students was the way they managed their involvement in the task rather than the way they processed the experience.

Participants in both groups required about five practice responses in order to acquire a new spelling response. On average, participants in the low performance group only required one more practice response in order to acquire a new spelling response.

The present data suggest that the number of practice responses is the critical variable for acquisition regardless of the teaching procedure used. This observation will be explored in Chapter 10.

REFERENCES

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition and maintenance of geography facts by elementary students with learning disabilities. *Journal of Behavioral Education, 3*, 217-233.

Drevno, G. E., Kimball, J. W., Possi, M. K., Heward, W. L., Gardner, R., & Barbetta, P. M. (1994). Effects of active student response during error correction on the acquisition, maintenance, and generalization of science vocabulary by elementary students: A systematic replication. *Journal of Applied Behavior Analysis, 27*, 179-180.

Johnson, P., Schuster, J., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6*, 437-458.

Nelly, E., & Smith, A. (2000). *PM Benchmark kit: An assessment resource for emergent - 12 years reading age*. Melbourne: Nelson ITP.

Nuthall, G. (2001). *The cultural myths and realities of teaching and learning.* Paper presented at the NZARE Annual Conference, Christchurch, New Zealand.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-406.

CHAPTER 10

DISCUSSION

The primary aim of the experiments described in this thesis was to explore variables affecting instructional efficiency. This final chapter reviews the believability of the results obtained in light of the experimental procedures employed, proposes a general model for analysing the factors affecting instructional efficiency, reviews the major findings of these experiments, identifies some of the main implications of the present experiments for teaching practice in the junior-school classroom, and suggests areas for future research.

PART ONE: REVIEW OF EXPERIMENTAL PROCEDURES

The believability of a set of experimental results depends upon the degree of control achieved over (a) the measurement of learning and (b) the experimental procedures used to measure the effects of experimental manipulations on rate of learning.

Computer Controlled Variables

It was decided to administer practice sessions using Macintosh LC 575 computers rather than human teachers and observers because a computer can provide a level of experimental control that is much greater than that which could be achieved by a human teacher. Karsh and Repp (1992, p. 465) state "the capacity of computer software to isolate and control specific instructional variables makes CAI a desirable context in which to examine these variables." Computer controlled variables in the present experiments included the size of the practice set, ordering of words for practice, the number of times that each word was presented, standardised stimulus displays, whether or not a prompt was presented within a trial, the form of the prompt, when the prompt was presented, whether the prompt disappeared prior to the learner responding or whether it stayed in view while the learner responded, whether or not feedback was provided after a response, the latency between the completion of the response and the presentation of feedback, the size, intensity and volume of the correct and error feedback, whether secondary responses were required, the latency between trials, the number of trials per session, and the number of responses per session. In addition to controlling these variables, the computers were also programmed to record participant's responses and response latencies and hence provided much higher levels of procedural reliability than could have been achieved by human recorders.

Experimenter Controlled Variables

Response opportunities. It is important to either control or record the number of practice opportunities in a learning experiment otherwise the experimenter cannot determine whether an observed effect is a function of (a) the independent variable, or (b) a difference in the number of response opportunities from one treatment to the next. The number of responses was controlled in the present experiments by ending each session once an equal number of predetermined response opportunities had occurred across treatments for each participant.

For the Error-Contingent Model Secondary-Response Treatment in Experiment 4 it was not possible to control both the number of responses and the number of trials. Previous studies (Barbetta, Heron, & Heward, 1993; Barbetta & Heward, 1993; Espin & Deno, 1989) investigating secondary-response opportunities controlled the number of trials. This had the advantage of the participant being presented with the instructional stimulus the same number of times. The disadvantage however was that the number of responses for the ASR Error-Correction Treatment was substantially higher than the Error-Correction Treatment. This created a confound in that the superior effectiveness of the ASR Error-Correction Treatment may have been due to the additional practice responses which occurred (Barbetta, Heron et al., 1993). In Experiment 4 this confound was controlled by controlling the total number of practice responses.

Participant learning histories. Participants were drawn from a pool of normally developing Year 2 children in a Decile 7 urban-primary school. Children were excluded from the pool of potential participants if they (a) had a developmental delay of 12 months or more below their chronological age, or (b) had a reading level 12 months or more below their chronological age, or (c) were unable to print all the letters of the alphabet, or (d) were unable to achieve a fluency level of 20 letters per minute on the 26 alphabet keys on a computer keyboard after 15 minutes of instruction. Experimental participants were selected at random from children who survived these screening tests. The same screening procedure was used for all four experiments. This limited but did not eliminate intersubject variability.

Experimental environment. The experiments were conducted within the school library which allowed the experimenter to control more extraneous stimuli than if the experiments were conducted within a classroom. The experiments were carried out in an area of the library away from other children, and from Experiment 2 onwards, screens were placed between computer stations to minimise interactions between participants and other children in the library.

Task difficulty across treatments. Task difficulty was controlled in Experiments 1 and 2 by assigning age-appropriate words randomly to treatments and additionally in Experiment 2 by removing difficult age-appropriate words from the pool of unknown words prior to randomisation. These procedures, however, did not satisfactorily control task difficulty. Task difficulty was better controlled in Experiments 3 and 4 by counting the number of letter pairs incorrect during pretesting and the total number of letter pairs in each word in the pool of unknown words. Unknown words were then assigned to either Treatment A or B to create two treatments of similar difficulty for each child. That is, two treatments in which the total number of letter pairs and the total number of letter pairs correct on the pretest were more closely matched.

Testing procedures. Testing procedures were well controlled by standardising the probe-testing procedure prior to the series of experiments and using the same procedure for each experiment.

Variables That Were Difficult to Control

Task difficulty across participants. Word difficulty across treatments was well controlled in Experiments 3 and 4 by matching the words assigned to each treatment according to the number of letter-pair errors on the pretest. However, it was very difficult to control the degree of word difficulty across participants. For example, a participant may have had 35 letter-pair errors out of a total 60 letter pairs in each treatment at the beginning of their experiment while another participant may have had 50 letter-pair errors out of a total of 78 letter pairs at the beginning of their experiment. In the present experiments it was not possible to produce word sets of similar difficulty across participants because (a) there were a relatively small and finite number of words at the 6- and 8-year old level of the spelling lists used for the present experiments, (b) different children made different errors (different letter-pair errors and different numbers of letter-pair errors) on the same word during pretesting, and (c) different participants acquired different words on different days of the experiment. The only possible solution to this problem is to use an invented task but artificial tasks lack social validity and were avoided for this reason.

Variations in levels of attention. Prior to Experiment 1 attempts were made to control attending responses by providing each participant with headphones to reduce extraneous stimuli. Despite this, variability in attending to extraneous stimuli was high as participants often talked to and distracted one another. Screens were placed between computer stations for Experiments 2, 3 and 4 in an attempt to reduce interaction between participants. While the screens stopped participants from seeing one another and levels of interaction were considerably lower, when interaction did occur it was more intense as participants needed to raise their voice in order to get one another's attention.

Variations in learning-to-learn skills. Participants were drawn from a pool of normally-developing children within the school. As expected, there were large variations in the achievement levels of individual children. During training it was observed that high-achieving children tended to (a) rehearse letters in the model of the correct spelling in both the antecedent and consequent position and (b) attend to the letter-pair error component in error-contingent models while lower achieving children tended not to engage in these learning-to-learn skills. As a result, it was necessary to teach some participants to attend to and rehearse the letters in the model. Despite this, the level of attention and self-rehearsal skills still varied greatly between participants throughout the experiments.

Variations in reactions to error feedback and error-contingent models. It was observed throughout the experiments that participants reacted differently to error feedback and the error-contingent models. Some participants attended closely to the error-contingent model while others either pressed the right arrow key or the NEXT WORD button almost immediately the correction was presented, apparently as an escape response.

Rate of Acquisition

Overall, the computer-administered spelling programme was effective for participants across all treatments. Participants acquired, on average, 12 new spelling responses after four 20-minute sessions. In Experiments 3 and 4, participants experienced two treatments per day and therefore acquired approximately 24 words after four school days with a total practice time of approximately two hours. This effectiveness compares well with previous research. In an invented spelling programme, Gettinger (1993) found that two high-ability 7-year olds acquired a mean of four words per week after three 15-minute sessions per week. Cunningham and Stanovich (1990) taught spelling to two groups of 24 6-year old children using different response modes and, after two hours of instruction, found that the children had acquired between 12 and 17 words. Berninger et al. (1998) found that 128 poor-spelling 6-year old children acquired a mean of 18 words after 24 sessions and a total time of 8 hours. Stevens and Schuster (1987) implemented a time-delay procedure with an 11-year old with spelling deficits. He acquired 15 words after 4.25 hours of instruction.

Several factors probably contributed to the effectiveness of the programme. First, the computer programme was easy for participants to learn to operate. Second, participants in this series of experiments actively responded to each instructional stimulus during instruction. Active responding has been shown to be more effective than passive responding (Narayan, Heward, & Gardner, 1990). Third, participants were provided with sufficient practice opportunities. Rosenshine and Stevens (1985, p. 386) state that teachers "should give the students enough practice that they become firm in their understanding and use of the new concepts or skills." Fourth, participants were provided with immediate feedback on every response. As Grant and Evans (1994, p 363) state, "the more frequent the feedback, the more effective the feedback will be in improving performance."

PART TWO: A MODEL OF THE ACQUISITION PROCESS FOR DISCRETE RESPONSES

The various events that can occur during the acquisition of a new response (such as a new spelling response) and the sequences of events that are possible within each learning interaction are shown in Figure 40.

Because new spelling responses are seldom acquired by 6- to 7-year old children following a single practice response, several practice responses are shown. Each practice response may or may not be preceded by some kind of prompt and correct responses may or may not be followed by some kind of feedback. Incorrect responses may or may not be

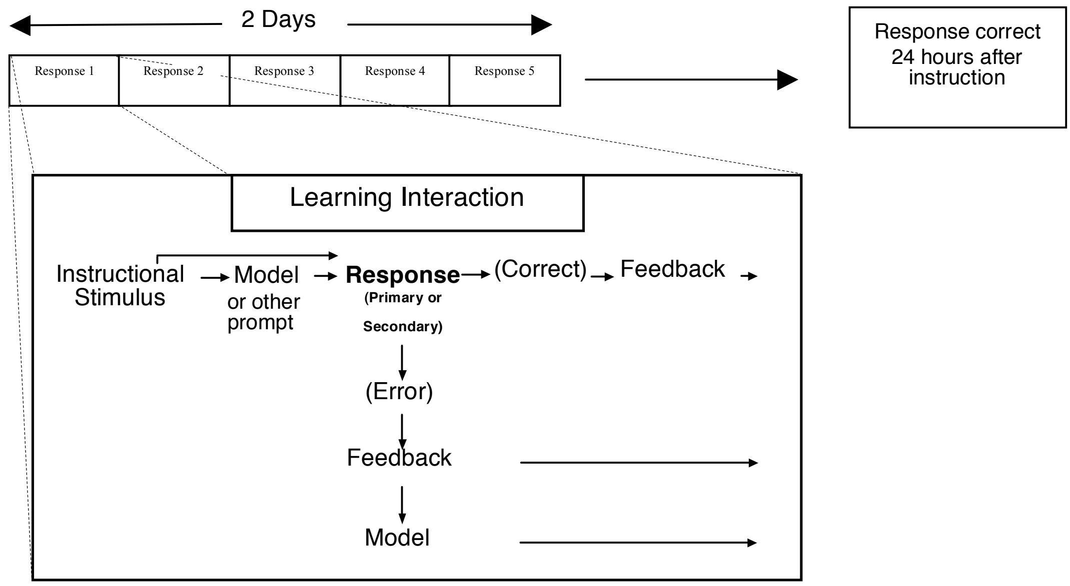


Figure 40. Diagram of the ways in which a response might be acquired.

followed by some kind of feedback, postcedent prompt, and/or secondary-response requirement.

Practice Response

During the course of a learning interaction a variety of response modes (e.g., oral, written, typed) are possible (Kearney & Drabman, 1993). Previous research has shown that children are able to learn new spelling responses using typing as the response mode (Berninger et al., 1998; Vaughn, Schumm, & Gordon, 1993). It is important that the set size is greater than the two or three items of information that young children can "hold in mind" at a time (Case, 1978). Otherwise the child may attend to recalling the response from stimuli other than the instructional stimulus.

Feedback

Feedback is widely considered to be an important variable in learning (Getsie, Langer, & Glass, 1985; Hattie, 1992; Lysakowski & Walberg, 1982; Walberg, 1991). Kulhavy, Yekovich and Dyer (1976, p. 522) suggest that "feedback is undoubtedly one of the most powerful tools in the arsenal of instructional design." However, variations in the use of the term feedback make it difficult to make comparisons across experiments. For example, Grimes (1981, p. 17) describes feedback as “information given to the learner about his or her performance” and then provides a description of 14 different kinds of feedback some of which include error corrections. Feedback in the present experiments was defined as a stimulus presented to a learner, contingent upon a response, which indicated whether the response was correct or incorrect. Error-contingent models, however, were treated as a separate variable.

The Antecedent Prompt

In the present experiments all prompts took the form of a model of the correct response. This visual model stayed in view until the child began to type thus allowing the learner to control the time over which the model was available for inspection. As soon as the child began to type, the model disappeared from the screen. It was decided to use a non-copying model because it has been shown that copying responses tend to result in little learning (Anderson, Kulhavy, & Andre, 1972; Kulhavy, 1977). That is, learners learn to copy the model rather than attending to and responding to critical dimensions of the instructional stimulus.

If an antecedent prompt is the only prompt present within the trial then this prompt provides the only opportunity for a transfer of stimulus control from the prompt to the instructional stimulus. Without an antecedent prompt it would be expected that a learner would respond incorrectly and, with no error-contingent prompt, no opportunity exists for the transfer of stimulus control from the prompt to the instructional stimulus.

It is good teaching practice to fade prompts gradually or at least not to fade a prompt until the learner can respond correctly without it. Fading of the model was not possible in the present experiments. In Experiments 1 and 2 the presentation or non-presentation of the antecedent model was dependent upon the accuracy level of the child’s practice responses and in Experiment 3 it was not possible to determine whether a correct response indicated acquisition or whether it simply indicated that the learner was responding correctly as a function of having just inspected the model.

The Error-Contingent Prompt

Correcting learner errors is considered to be important for acquisition (Barringer & Gholson, 1979; Heubusch & Lloyd, 1998; Kulhavy, 1977; Rosenshine, 1983). A variety of error-correction procedures have been described. For example, error-correction procedures have included the presentation of a stimulus indicating an error response (feedback) and a model of the correct response while in others it has consisted of feedback and a weak or strong prompt for a secondary-response opportunity (Browder & Xin, 1998; Heubusch & Lloyd, 1998; Johnson, Schuster, & Bell, 1996). Feedback following errors is often referred to as error correction (Alvarado-Gomez & Belfiore, 2000; Gettinger, 1993a; Grskovic & Belfiore, 1996). These variations in error-correction procedures make it difficult to compare results across experiments. In the present series of experiments the error-correction procedure was standardised by presenting feedback (as described above) immediately followed by a model of the correct response in the same form and intensity following every incorrect response.

An error-contingent prompt is critical if it is the only prompt present within the trial as it provides the only opportunity for the transfer of stimulus control from the prompt to the instructional stimulus. Errors are therefore learning opportunities under error-contingent model treatments provided that the child receives another opportunity to respond to the instructional stimulus.

Secondary Response Opportunity

A secondary-response opportunity occurs when a student emits a second response to the instructional stimulus within the same trial following an error response, and usually occurs following some form of error correction (Barbetta, Heron et al., 1993). Further responses may occur within a trial until the student emits the correct response (e.g., Baker, 1992). The effects of requiring a secondary response were only explored in Experiment 4. Secondary response opportunities were not provided following errors in Experiments 1, 2 and 3 in order to control the number of practice responses in these experiments.

Number of Practice Responses Required for Acquisition

Participants in the present series of experiments required an average of five practice responses to acquire a new spelling response across all treatments. This appeared to be critical for acquisition because the number of practice responses required was very stable regardless of the achievement level of the participant, the accuracy level during instruction, whether the model was in the antecedent or consequent position, or whether participants engaged in secondary responses.

Distribution of Practice Responses in Time

Practice responses are distributed in time. Practice responses may occur within a single session or distributed over a series of sessions. The sessions might be within a single day, distributed over several days, or even weeks apart. In the present series of experiments almost all learned responses were acquired within two sessions distributed over a two-day period.

PART THREE: FACTORS AFFECTING INSTRUCTIONAL EFFECTIVENESS AND EFFICIENCY

The experiments described in this thesis attempted to measure the effects of a number of variables on rate of acquisition and instructional efficiency in 6- to 7-year old children. Experiments 1 and 2 examined the relationship between level of correct responding during instruction and rate of learning in a spelling task and found that, under antecedent-model conditions, the rate of acquisition was greater with prompting ratios which produced a greater proportion of correct responses during instruction. Experiment 3 compared the effects of presenting a model of the correct response (a) prior to responding and (b) following incorrect responses and found that rate of acquisition was only slightly higher under antecedent-model conditions. Both procedures were, however, equally efficient in terms of responses to criterion. Experiment 4 studied the effects of presenting an error-contingent model followed by either (a) the next trial or (b) a secondary response requirement and found that moving to the next trial was slightly more effective but that requiring a secondary response was slightly more efficient in terms of the number of responses to criterion.

Practice Response Variables

The mode of responding in the present series of experiments was typing. Participants were trained to type at a rate of at least 20 letters per minute prior to the experiment. This level of typing fluency was sufficient given that the learners were practising within the acquisition phase. Informal observations did not reveal any participants who were frustrated at finding letter keys. A higher level of typing fluency would, however, have been required had the experiment continued into the fluency building phase of learning.

For younger learners Engelmann and Carnine (1991) suggest a practice set size of six items. In the present experiments the practice set size was 10 words and this was constant across experiments. The decision to use a practice set size of 10 words was made in order to avoid a ceiling effect where participants might acquire all the words in a set in a single session. Ceiling effects would have meant that the measurement of effectiveness and efficiency would have been compromised.

Feedback

In the present investigations feedback was standardised across all experiments (a) by presenting it immediately after each response and (b) by controlling the intensity of the audio and the visual dimensions of feedback following both correct and incorrect responses.

Feedback was presented immediately following each response to standardise the latency between each response and the feedback. Because feedback was controlled in this way the present experiments provide no data regarding the effects of feedback or its role as a determinant of instructional efficiency.

Some children found the error feedback and the presentation of the error-contingent model aversive. For these children it is interesting to ask which of the error-contingent events they were trying to escape from. In Experiments 3 and 4 the experimenter often observed participants (a) complain when they responded incorrectly, or (b) turn away from the screen or, (c) press the right arrow key or click the NEXT WORD button immediately following the presentation of error feedback and an error-contingent model. These reactions from participants were also observed in Experiment 2 where errors were followed only by feedback and no error correction. It therefore appears that participants who reacted negatively were reacting negatively to the error feedback.

The Antecedent Prompt

Participants in the Antecedent-Model Treatment in Experiment 3, on average, acquired 12.0 words after four sessions, and required 5.5 practice responses on these words to learn these words. It was initially thought that a non-copying model would provide a high level of stimulus control over the correct response. However, in the antecedent-model treatments in Experiments 1, 2 and 3, the antecedent model resulted in a correct response on only 70% of the first round practice trials in each session. The failure of the model to control the correct response on the remaining occasions may have been due to a number of variables. First, the short-term memory capacity of 6- to 7-year old children only allows them to recall two to three items (Case, 1978). This may have been insufficient for children who needed to remember more than one or two letter pairs. Second, the experimenter frequently observed individual children becoming distracted by other children while in the process of typing their response.

The Error-Contingent Prompt

Participants in the error-contingent model treatments in Experiments 3 and 4, on average, acquired 10.4 words after four sessions and required a mean of 5.5 practice responses on these words in order to acquire them. The number of responses to criterion was the same as the Antecedent-Model Treatment in Experiment 3.The error-contingent model used in Experiments 3 and 4 exercised stimulus control over the correct response on only 26% of subsequent practice responses (within a session) involving words previously spelled incorrectly. This is probably due to the fact that the ten words in each session were randomly re-ordered after every 10 practice responses. This meant (a) that the number of trials interpolated between the error-contingent model and the next trial on that word could range from 1 to 18 trials and (b) that the response delay following an error-contingent model could range from approximately 40 seconds to approximately 14 minutes. This delay will have greatly reduced the probability of the error-contingent model functioning as a prompt for the next unprompted presentation of that word.

Under the error-contingent model treatments in Experiments 3 and 4 transfer of stimulus control (that is, acquisition) was probably demonstrated once a participant responded correctly during practice because transfer of stimulus control could only occur from the model presented in the consequent position on a previous trial. After two correct responses during practice under these conditions participants almost always responded correctly 24 hours later.

Secondary-response opportunity. In the Error-Contingent Model Secondary-Response Treatment in Experiment 4 participants acquired an average 10.4 words and required an average of 4.6 practice responses on these words in order to acquire them. It will be recalled from Chapter 9 that an antecedent model in Experiment 3 prompted a correct primary response on 70% of occasions whereas an error-contingent model prompted a correct secondary response on only 56% of occasions even though the latency between the model and the response was very small in both cases. A possible reason for the lower effectiveness of the prompt for the secondary response was the fact that some of the children found the error feedback aversive and therefore attended less to the model in the error-contingent position.

Effects of consequences. Consequences exert a powerful effect on behaviour (Catania, 1998; B. F. Skinner, 1953). One type of consequence is feedback and this too is thought to have a powerful effect on learning and achievement (Hattie, 1992; Kulik & Kulik, 1988; Lysakowski & Walberg, 1982; Walberg, 1991). In the present series of experiments, however, variations in consequences had only a small effect on rate of acquisition. This was true whether a correction (a model) was presented in the error consequence position or not, and whether a secondary response was or was not required. It might be that similar rates of acquisition were observed because in every trial in each experiment (a) feedback was standardised and presented and (b), apart from a few trials in Experiment 2, an opportunity for transfer of stimulus control was present in the form of either an antecedent prompt or a postcedent prompt.

Number of Practice Responses Required for Acquisition

It will be recalled from Chapter 9 that participants across all experiments required, on average, five practice responses per word in order to acquire a new spelling response (respond correctly 24 hours following instruction). Few experiments have produced such consistent results as those found in this series of experiments. It appears that a sufficient number of practice responses distributed in time is the critical variable necessary to acquire a new spelling response provided an opportunity for transfer of stimulus control from the prompt to the practice stimulus is presented with each practice opportunity. Variability in the results from previous experiments may be due to variability in the way in which acquisition is defined, the type of learner, and type of learning task.

Definition of acquisition. Different investigators have used different definitions of acquisition. When acquisition was defined as responding correctly immediately following instruction Dineen, Clark and Risley (1977) and Okyere, Heron and Goddard (1997) found that 9- to 10-year old children required a mean of seven and four trials respectively to acquire a spelling response. When acquisition was defined as responding correctly 24 hours following instruction Axelrod, Kramer, Appleton, Rockett, and Hamlet (1984) found that a normally developing 11-year old participant required about six trials to acquire a spelling response.

Type of learner. Nulman and Gerber (1984) found that an 8-year old with a developmental delay required about eight trials to acquire a spelling response. Gerber (1984) found that a 10-year old with a developmental delay required about six trials, on average, to acquire a new spelling response. When the task was acquiring sight words, Wolery, Ault, Gast, Doyle and Mills (1990) found that 7- to 8-year olds classified as developmentally delayed required about 25 trials per word to acquire sight-word responses when acquisition was defined as responding correctly immediately following instruction. Barbetta, Heward and Bradley (1993) found that 8- to 9-year olds classified as developmentally delayed required an average of about 14 trials per word when sight-word responses were defined as responding correctly 24 hours following instruction. In acquiring spelling responses (when acquisition was measured during or immediately following instruction) normally developing 7-year olds (Cates et al., 2003) required about four trials per word while normally developing 9-year olds (Okyere et al., 1997) also required about four trials per word.

Learning task. Ten-year olds classified as developmentally delayed required about five trials to acquire health facts when acquisition was measured a few hours after instruction (Sterling, Barbetta, Heward, & Heron, 1997). This compares to 9- to 12-year olds classified as developmentally delayed who required about 12 trials to learn maths facts when acquisition was measured immediately following instruction (Koscinski & Gast, 1993).

Apart from Wolery et al. (1990), participants in the above experiments required between 4 and 15 trials in order to acquire a discrete academic response. The number of trials in the Wolery et al. experiment is higher than other experiments because Wolery et al. reported the number of practice responses (primary and secondary), not the number of trials. All other experiments either failed to control or failed to report the number of response opportunities within the experiment. For example, participants in Cates et al. (2003) required four learning trials to acquire a spelling word. However, this experiment used an overcorrection procedure (writing the word correctly three times contingent upon an error). This means the total number of practice responses was much greater than four per word. Therefore, while the above experiments provide information regarding the number of learning trials required in order to acquire a typical classroom-type response under various conditions they tell us little about the number of practice responses required. Failure to control the number of practice responses is not limited to the above experiments. It is an extremely common weakness in applied studies investigating rates of acquisition (e.g., Morton, Heward, & Alber, 1998; Okyere et al., 1997; Schermerhorn & McLaughlin, 1997; Wirtz, Gardner, Weber, & Bullara, 1996). It will be recalled from the Chapter 6 review of the research into the correction of learner errors that none of the 36 studies in the review controlled the number of practice opportunities. This makes it impossible to tell whether differences in effectiveness are the result of the independent variable or the number of uncontrolled practice responses. If the number of practice responses is the critical variable for acquisition, and this variable is uncontrolled in almost all of the teaching research, then the results from almost all of the teaching research may be uninterpretable.

Distribution of Practice Responses in Time

As the number of practice responses per day was usually three it meant that the five practice responses that were required in order to acquire a word were distributed over two days. However, it cannot be ascertained whether it was (a) five practice responses, or (b) the distribution of practice responses over two days, or (c) five practice responses distributed over two days that was critical in acquiring a word.

It is almost certainly the case that the distribution of practice responses in time is important for acquisition. This raises the question of how much time should elapse between practice responses for responses to be learned and remembered. While it has been repeatedly found that practice which is spaced in time produces faster rates of acquisition than practice that is massed (Dempster & Farris, 1990), there is little research evidence regarding the spacing of specific response opportunities, that is, the interresponse interval. A response practised, say, five times in a row is likely to be forgotten. Likewise, if the five practice responses are widely spaced (e.g., one practice response per week) then it is also likely that the response will not be remembered.

Nuthall (2000) investigated conditions where responses were learned and remembered and conditions where responses were learned and forgotten. Nuthall's research involved developing tests based on a sample of possible learning outcomes for to-be-taught science and social studies units for 9- to 12-year old children. These tests were administered to a sample of three to four students prior to the unit. During the unit, observers continuously recorded these students' behaviours using video cameras and microphones, and recorded all interactions with lesson-relevant content. Records were made each day of everything each student read, wrote and saw. An outcome test and interview was administered about two to three weeks after the unit. In some studies long-term outcome tests and interviews were also carried out.

Based on the results from several studies (e.g., Nuthall, 1999b; Nuthall & Alton-Lee, 1993), Nuthall (2000, p. 95) concluded that when students engage with concept-relevant information, they construct a representation of that information. This information has a life of two days and if "a single representation is not connected to a further related representation within the 2 days, it disappears from the working memory." Although yet to be demonstrated experimentally, Nuthall concluded the new concepts that the student worked on at least three to four times with no more than two days in between any two experiences were learned and remembered while other new concepts were not. These results suggest that a series of learning opportunities will result in acquisition only if they occur with no more than two days between any pair of response opportunities. Practice responses in the present series of experiments were distributed one day apart and this may have been a variable in the acquisition of new spelling responses. This possibility clearly warrants further experimental analysis.

PART FOUR: IMPLICATIONS

Classroom Practice

It is clear from the results of the present experiments that the instructional conditions provided by the programmed spelling activities were effective and efficient in helping 6- and 7-year old children to acquire new spelling responses. On average, participants learned the correct spelling of 12 new words during the course of four 20-minute sessions. The instructional conditions provided during these experiments included daily practice, individualised word lists, control over task difficulty, non-copying models of the correct spelling, immediate feedback following every response, and practice sets in which words were removed once they had been acquired and replaced with new words.

However, the level of control over practice provided in the present experiments cannot be provided in the classroom setting where one teacher is responsible for the learning activities of 25 to 35 children. This raises questions regarding the generalisability of the present findings to instructional settings such as classrooms.

It is possible to emulate many of the experimental conditions of the present experiments in classrooms. For example, children can be taught to use the cover, copy, and compare (CCC) rehearsal procedure (Murphy, Hern, Williams, & McLaughlin, 1990; C. H. Skinner, McLaughlin, & Logan, 1997) when studying new spelling words. The CCC procedure requires students to attend to the model of the spelling word, cover it, write the word, and then compare the response with the model. They then either have a secondary-response opportunity or move on to the next trial. Provided that children follow the procedure and have the prerequisite social skills, this emulates the non-copying prompt, feedback and presentation of the model after each response.

Children can be organised to practise new material (such as spelling words) in pairs in a class wide peer-tutoring programme (Kohler & Greenwood, 1990; Sideridis et al., 1997). Children can be assigned to pairs based on their current spelling or reading level so that both children in a pair are learning words at the same level of difficulty. This also has the advantage that each child, while not necessarily being able to spell each word, would be able to read the words that their partner is learning to spell.

The spelling programme could begin with each child testing the other to find an individual pool of unknown words from words provided by the teacher at the appropriate level. Only the teacher however could establish the difficulty level at which each child would be working. Unknown words could then be assigned to a spelling set of 6 to 10 words depending upon the age of the child. Provided the teacher scheduled a daily spelling practice time, daily sessions (apart from the first) might begin with each child testing each other on their individual words practised 24 hours earlier. Each correctly spelled word would be removed from the practice set once it had been acquired and replaced with a word from the pool of unknown words to keep the set size the same. After testing, each child could practise their spelling list three times using the CCC procedure. Provided each child is adequately trained in how to work as a partner, paired practice can provide the supervision over word selection, CCC practice and acquisition testing in a way in which a teacher is unable to do individually for 30 children simultaneously (Gordon, Vaughn, & Schumm, 1993).

Current (traditional) spelling programmes are often not very effective or efficient (Brown, 1990). In a traditional approach, students are given a list of, say, 10 words on Monday, the words are written in sentences on Tuesday, they practise word patterns with the words on Wednesday, write a story with all the spelling words on Thursday, and have a test on the words on Friday. Regardless of the performance on the Friday test, students are presented with a new list of spelling words the following week, and the cycle is repeated (Heron, Okyere, & Miller, 1991).

This approach appears to be rather ineffective and is certainly inefficient for several reasons. First, it is possible that a large proportion of the child's practice responses are copying responses. For example, a child writing a spelling word in a sentence may simply copy the spelling word paying minimal attention to the spelling of it so they can continue to complete the sentence. Second, it is possible that participants may only generate one practice response per day. For example, writing the spelling word in a story may be the only practice response of that word within the spelling session. Third, feedback may not be present and, if it is, may be delayed until the end of the session. For example, after a child has written a spelling word in a story, the child may continue to write the story without comparing the spelling of the word with a model of the word. Alternatively, the child may check all words at the completion of the story (if time permits), and may or may not correct any incorrect spellings. Fourth, the point at which a new spelling response is acquired cannot be ascertained. For example, a word might be acquired during the Tuesday spelling session but it continues to be practised taking away valuable practice time from learning other spelling words. Fifth, a new set of words is presented on Monday regardless of performance on the previous Friday test. This means that non-acquired words that have been practised for four days are no longer practised.

Dissatisfaction with the traditional weekly list approach has led many teachers to abandon formal spelling activities in favour of a whole language approach to reading, writing and spelling. A key assumption of this approach is that children will learn to spell by engaging in reading and writing. However, Graham (2000, p. 244) argues that "incidental learning from reading and writing cannot account for most gains in spelling." The authentic writing approach suffers from the same weaknesses as the traditional spelling approach. That is, the number of practice responses may not be enough to acquire a word, feedback may not be present, and immediate feedback is rare. In addition, specific spelling responses may be practised, say, once in the first week of the school term, none in the second, twice in the third week etc. This means that many more practice responses are required in order to acquire the spelling response than would be the case if the word was practised on a daily basis. A variation of the authentic approach encourages students to use invented spellings. This is based on the assumption that children's spelling will improve as their reading and writing improves. Given that no models of correct spelling are provided, no practice opportunities for the spelling response occur and therefore no feedback is provided, it is not surprising that Brown (1990) concludes there is no evidence that an invented spelling programme improves spelling ability.

It seems likely that implementation of a peer-tutoring procedure like the one described above could result in faster rates of acquisition than is presently the case. However, there is no simple solution as changing just one feature will probably not improve either effectiveness or efficiency. For a sizeable increase in efficiency, improved individualisation, prompting, practice and feedback will all need to be achieved.

PART FIVE: FUTURE RESEARCH

The present experiments identified a number of unanswered questions. One of the most important questions was why the experimental changes to error consequences had so little effect on rates of acquisition and instructional efficiency in the present experiments. Across all experiments, participants acquired, on average, about 12 new spelling responses after four sessions and required about five practice responses to acquire each new spelling response. The results from the present series of experiments suggest that provided the correct spelling is prompted and feedback follows practice responses it does not matter too much whether the prompt appears in either the antecedent or consequent position for either a primary or secondary response. What seems to be critical is that there is an opportunity for transfer of stimulus control from the prompt to the orally stated word. Future research could investigate this by isolating and controlling variables within the transfer of stimulus control opportunity, and observing the effects of the different forms of prompting on rate of acquisition and instructional efficiency. This may identify the variables that must be present in order for a transfer of stimulus control to occur.

In the present experiments new spelling responses tended to be acquired over a two-day period. Nuthall (1999a; 2000) has found that the time that elapses between each learning experience plays a significant role, with a gap of more than two days resulting in the failure of relevant experiences to result in learning. The scheduling of practice was not manipulated during the present experiments so it could not be ascertained whether a 24-hour distribution of practice responses affected the rate of acquisition that was observed. All that is known is that participants were provided with three practice responses per word per session, and required five practice responses in order to acquire a spelling response, and that this tended to occur after two days. Nuthall's results suggest that the way in which learning opportunities are distributed in time may be a critical determinant of acquisition. Future research could investigate the effects of varying interresponse times by conducting experiments in which the number of practice responses is controlled and the distribution of these responses is varied in time.

The present experiments found that the amount of practice required in order to acquire a new spelling response was very consistent across participants. However, the instructional conditions which are required differ depending upon the type of learning outcome sought (Engelmann & Carnine, 1991). It is therefore likely that the amount of practice required in order to acquire different types of responses will differ (e.g., Koscinski & Gast, 1993; Sterling et al., 1997). Future research might therefore investigate the effects of different tasks on rate of acquisition perhaps by implementing computer-controlled instruction using a variety of different types of classroom-like tasks.

The participants in the present series of experiments were 6- and 7-year old children who required about five practice responses to acquire a spelling response. Participants in Nuthall's studies were 9- to 12-year old children who required only three to four response opportunities in order to acquire social studies and science type responses. It would be interesting to know whether older children require fewer response opportunities for acquisition (and if so why) or whether the apparently more rapid learning in the Nuthall studies is simply a function of the different learning tasks or the less stringent acquisition measures used in Nuthall's studies.

It will be recalled that the decision to use a practice set size of 10 words in the present series of experiments was an experimental decision. It is likely however that the set size affects rate of acquisition. Gettinger, Bryant and Fayne (1982) found that a set size of three spelling words a day was more effective than set sizes greater than three spelling words with students classified as learning disabled while Engelmann and Carnine (1991) suggest a practice set size of six items for normally-developing younger children. Future research might investigate the effects of variations in set size on rate of acquisition and instructional efficiency. This could easily be achieved by using the present computer programme and experimental procedures.

It was thought that a correct response during practice under error-contingent model treatments might have been a measure of acquisition because unprompted correct practice responses were very often correct 24 hours later. Future research might explore this possibly by ceasing further practice on a particular response following the first unprompted correct response and testing this response for recall 24 hours later. If it is found that unprompted correct responses are correct 24 hours later then this has important implications for classroom instruction.

In Experiment 3 it was found that an incorrect practice response on the final round of a session under the Error-Contingent Model Treatment tended to be incorrect on the 24-hour probe test. This suggests that an error-contingent model in this position has little effect on acquisition but this needs to be checked by experimental analysis.

CONCLUSION

The experiments described in this thesis investigated variables that affect instructional efficiency by employing specifically programmed computers to manage and control instructional variables within each experiment. It was found that 6- to 7-year old children, on average, acquired about 12 new spelling responses after four 20-minute sessions involving 3 practice responses on each of 10 unknown spelling words. Each newly acquired spelling response required about five practice responses and was acquired over a two-day period. The number of practice responses required to acquire a spelling response was very stable across participants in all treatments despite variations in accuracy levels during instruction, whether the prompt occurred in the antecedent or postcedent position and whether or not a secondary response was required following error corrections. These results suggest that an opportunity for the transfer of stimulus control from the prompt (model of the correct spelling) to the practice stimulus (the spoken word) is more critical for acquisition than where within the trial this occurs. It was also thought that the distribution of acquired responses over a two-day period may have been important for acquisition. This however was not experimentally investigated.

Variations in rates of acquisition amongst participants were observed. These appear to have been the result of variations in participant's entry skills, levels of attending, and task-management skills. Although rates of acquisition differed between high-achieving participants and low-achieving children, there was little difference in the number of practice responses required for acquisition between these two groups.

It was observed that most of the 6- to 7-year old participants found error feedback aversive and this appeared to result in reduced attention to models of the correct spelling when these occurred following errors.

The results suggest that children could acquire many more spelling words in a classroom programme than is currently the case with either traditional or whole language approaches to the teaching of spelling. For example, a class wide peer-tutoring programme using a procedure such as the cover, copy, compare rehearsal procedure on a daily basis could emulate the conditions provided during the present experiments.

Skinner (1968) argued that it was the feedback following correct responses that was critical to acquisition. He argued that a learning trial should contain a prompt that generates a correct response, and then feedback to reinforce the correct response. Kulhavy's (1977) view, however, was that feedback's main effect is to correct errors. According to Kulhavy, errors interact with response certitude. Response certitude is the degree of confidence that the learner has that their response is correct. Kulhavy argued that feedback has its strongest effect on acquisition when response certitude is high but the response is incorrect. That is, feedback is most effective when the learner believes the response is correct but is in fact an error. Kulhavy (1977, p. 211), however, uses the term feedback "in a generic sense to describe any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong." Therefore, it may be that error feedback is most effective when it takes the form of a correction (which informs the learner of the correct response) and, therefore provides an opportunity for the transfer of stimulus control to occur. When response certitude is high the learner attends more closely to the error correction and hence is more likely to respond correctly the next time that this response is required.

The results from the present series of experiments have shown that rates of acquisition are similar whenever the opportunity for a transfer of stimulus control (either in the antecedent or consequent position) occurs. This result suggests that it is probably not the feedback per se, or the learner's attention to this feedback which is critical but rather the opportunity for a transfer of stimulus control to occur. It may be that neither responding correctly with feedback, as Skinner argued, nor responding incorrectly with feedback, as Kulhavy argued, are the critical variables. Rather, it seems to be the number of occasions on which it is possible for a transfer of stimulus control to occur which is critical for acquisition.

It was suggested that future research could investigate (a) the variables which are necessary for the transfer of stimulus control, (b) the generality of the observation that children require five practice responses in order to acquire discrete academic responses, and (c) the effects on rates of acquisition and instructional efficiency of varying the distribution in time of practice responses for children who are learning various types of academic skills.

REFERENCES

Alvarado-Gomez, M., & Belfiore, P. J. (2000). Effects of error correction on Spanish spelling words for students whose primary language is Spanish. *Journal of Behavioral Education, 10*, 213-221.

Anderson, R. C., Kulhavy, R., & Andre, T. (1972). Conditions under which feedback facilitates learning from a programmed lesson. *Journal of Educational Psychology, 63*, 186-188.

Axelrod, S., Kramer, A., Appleton, E., Rockett, T., & Hamlet, C. C. (1984). An analysis of the relevance of topographical similarity on positive practice of spelling errors. *Child and Family Behavior Therapy, 6*, 19-31.

Baker, J. D. (1992). Correcting the oral reading errors of a beginning reader. *Journal of Behavioral Education, 2*, 337-343.

Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 111-119.

Barbetta, P. M., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition and maintenance of geography facts by elementary students with learning disabilities. *Journal of Behavioral Education, 3*, 217-233.

Barbetta, P. M., Heward, W. L., & Bradley, D. M. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110.

Barringer, C., & Gholson, B. (1979). Effects of type and combination of feedback upon conceptual learning by children: Implications for research in academic learning. *Review of Educational Research, 49*, 459-478.

Berninger, V., Abbott, R. D., Brooks, A., Abbott, S. P., Rogan, L., Reed, E., et al. (1998). Early intervention for spelling problems: Teaching functional spelling units of varying size with a multiple-connections framework. *Journal of Educational Psychology, 90*, 587-605.

Berninger, V., Abbott, R. D., Rogan, L., Reed, E., Abbott, S. P., Brooks, A., et al. (1998). Teaching spelling to children with specific learning disabilities: The mind's ear and eye beat the computer or pencil. *Learning Disability Quarterly, 21*, 106-122.

Browder, D. M., & Xin, Y. P. (1998). A meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education, 32*, 130-153.

Brown, A. S. (1990). A review of recent research on spelling. *Educational Psychology Review, 2*, 365-397.

Case, R. (1978). A developmentally based theory and technology of instruction. *Review of Educational Research, 48*, 439-463.

Catania, A. C. (1998). *Learning (4th ed.)*. New Jersey: Prentice Hall.

Cates, G. L., Skinner, C. H., Watson, T. S., Meadows, T. J., Weaver, A., & Jackson, B. (2003). Instructional effectiveness and instructional efficiency as considerations for data-based decision making: An evaluation of interspersing procedures. *School Psychology Review, 32*, 601-616.

Cunningham, A. E., & Stanovich, K. E. (1990). Early spelling acquisition: Writing beats the computer. *Journal of Educational Psychology, 82*, 159-162.

Dempster, F. N., & Farris, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23*, 97-101.

Dineen, J. P., Clark, H. B., & Risley, T. R. (1977). Peer tutoring among elementary students: Educational benefits to the tutor. *Journal of Applied Behavior Analysis, 10*, 231-238.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and application*. Eugene: ADI Press.

Espin, C. A., & Deno, S. L. (1989). The effects of modeling and prompting feedback strategies on sight word reading of students labeled learning disabled. *Education and Treatment of Children, 12*, 219-231.

Gerber, M. M. (1984). Techniques to teach generalizable spelling skills. *Academic Therapy, 20*, 49-58.

Getsie, R. L., Langer, P., & Glass, G. V. (1985). Meta-analysis of the effects of type and combination of feedback on children's discrimination learning. *Review of Educational Research, 55*, 9-22.

Gettinger, M. (1993). Effects of invented spelling and direct instruction on spelling performance of second-grade boys. *Journal of Applied Behavior Analysis, 26*, 281-291.

Gettinger, M. (1993a). Effects of error correction on third graders' spelling. *Journal of Educational Research, 87*, 39-45.

Gettinger, M., Bryant, N. D., & Fayne, H. R. (1982). Designing spelling instruction for learning disabled children: An emphasis on unit size, distributed practice, and training for transfer. *The Journal of Special Education, 16*, 439-448.

Gordon, J., Vaughn, S., & Schumm, J. S. (1993). Spellling interventions: A review of literature and implications of instruction for students with learning disabilities. *Learning Disabilities Research and Practice, 8*, 175-181.

Graham, S. (2000). Should the natural learning approach replace spelling instruction? *Journal of Educational Psychology, 92*, 235-247.

Grant, L., & Evans, A. (1994). *Principles of behavior analysis*. New York: HarperCollins.

Grimes, L. (1981). Error analysis and error correction procedures. *Teaching Exceptional Children, 14*, 17-21.

Grskovic, J. A., & Belfiore, P. J. (1996). Improving the spelling performance of students with disabilities. *Journal of Behavioral Education, 6*, 343-354.

Hattie, J. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*, 5-13.

Heron, T. E., Okyere, B. A., & Miller, A. D. (1991). A taxonomy of approaches to teach spelling. *Journal of Behavioral Education, 1*, 117-130.

Heubusch, J. D., & Lloyd, J. W. (1998). Corrective feedback in oral reading. *Journal of Behavioral Education*, 63-79.

Johnson, P., Schuster, J., & Bell, J. K. (1996). Comparison of simultaneous prompting with and without error correction in teaching science vocabulary words to high school students with mild disabilities. *Journal of Behavioral Education, 6*, 437-458.

Karsh, K. G., & Repp, A. C. (1992). Computer-assisted instruction: Potential and reality. In N. N. Singh & I. L. Beale (Eds.), *Learning disabilities: Nature, theory and treatment*. New York: Springer-Verlag.

Kearney, C. A., & Drabman, R. S. (1993). The write-say method for improving spelling accuracy in children with learning disabilities. *Journal of Learning Disabilities, 26*, 52-56.

Kohler, F. W., & Greenwood, C. R. (1990). Effects of collateral peer supportive behaviors within the classwide peer tutoring program. *Journal of Applied Behavior Analysis, 23*, 307-322.

Koscinski, S. T., & Gast, D. L. (1993). Use of constant time delay in teaching multiplication facts to students with learning disabilities. *Journal of Learning Disabilities, 26*, 533-544.

Kulhavy, R. (1977). Feedback in written instructions. *Review of Educational Research, 47*, 211-232.

Kulhavy, R., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*, 522-528.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79-97.

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559-578.

Morton, W. L., Heward, W. L., & Alber, S. R. (1998). When to self-correct? A comparison of two procedures on spelling performance. *Journal of Behavioral Education, 8*, 321-335.

Murphy, J. F., Hern, C. L., Williams, R. C., & McLaughlin, T. F. (1990). The effects of CCC approach in increasing spelling accuracy with learning disabled students. *Contemporary Educational Psychology, 15*, 378-386.

Narayan, J. S., Heward, W. L., & Gardner, R. (1990). Using response cards to increase student participation in an elementary classroom. *Journal of Applied Behavior Analysis, 23*, 483-490.

Nulman, J. R., & Gerber, M. M. (1984). Improving spelling performance by imitating a child's errors. *Journal of Learning Disabilities, 17*, 328-333.

Nuthall, G. (1999a). Learning how to learn: The evolution of students' minds through the social processes and culture of the classroom. *International Journal of Educational Research, 31*, 141-256.

Nuthall, G. (1999b). The way students learn: Acquiring knowledge from an integrated science and social studies unit. *Elementary School Journal, 99*, 303-341.

Nuthall, G. (2000). The role of memory in the acquisition and retention of knowledge in science and social studies units. *Cognition and Instruction, 18*, 83-139.

Nuthall, G., & Alton-Lee, A. (1993). Predicting learning from student experience of teaching: A theory of student knowledge construction in classrooms. *American Educational Research Journal, 30*, 799-840.

Okyere, B. A., Heron, T. E., & Goddard, Y. (1997). Effects of self-correction on the acquisition, maintenance, and generalization of the written spelling of elementary school children. *Journal of Behavioral Education, 7*, 51-69.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal, 83*, 335-351.

Rosenshine, B., & Stevens, R. (1985). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching (3rd. ed)* (pp. 376-391). New York: Macmillan Publishing Company.

Schermerhorn, P. K., & McLaughlin, T. F. (1997). Effects of the Add-A-Word spelling program on test accuracy, grades, and retention of spelling words with fifth and sixth grade regular education students. *Child and Family Behavior Therapy, 19*, 23-35.

Sideridis, G. D., Utley, C., Greenwood, C. R., Delquadri, J. C., Dawson, H., Palmer, P., et al. (1997). Classwide peer tutoring: Effects on the spelling performance and social interactions of students with mild disabilities and theier typical peers in an integrated instructional setting. *Journal of Behavioral Education, 7*, 435-462.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Skinner, B. F. (1968). *The technology of teaching*. New York: Meredith Corporation.

Skinner, C. H., McLaughlin, T. F., & Logan, P. (1997). Cover, copy, and compare: A self-managed academic intervention effective across skills, students, and settings. *Journal of Behavioral Education, 7*, 295-306.

Sterling, R. M., Barbetta, P. M., Heward, W. L., & Heron, T. E. (1997). A comparison of active student response and on-task instruction on the acquisition and maintenance of health facts by fourth grade special education students. *Journal of Behavioral Education, 7*, 151-165.

Stevens, K. B., & Schuster, J. W. (1987). Effects of a constant time delay procedure on the written spelling performance of a learning disabled student. *Learning Disability Quarterly, 10*, 9-16.

Vaughn, S., Schumm, J. S., & Gordon, J. (1993). Which motoric condition is most effective for teaching spelling to students with and without learning disabilities. *Journal of Learning Disabilities, 26*, 191-198.

Walberg, H. J. (1991). Productive teaching and instruction: Assessing the knowledge base. In H. C. Waxman & H. J. Walberg (Eds.), *Effective Teaching: Current research*. Berkeley: McCutchan Publishing Corporation.

Wirtz, C. L., Gardner, R., Weber, K., & Bullara, D. (1996). Using self-correction to improve the spelling performance of low-achieving third graders. *Remedial and Special Education, 17*, 48-58.

Wolery, M., Ault, M. J., Gast, D. L., Doyle, P. M., & Mills, B. M. (1990). Use of choral and individual attentional responses with constant time delay when teaching sight word reading. *Remedial and Special Education, 11*, 47-58.

APPENDIX

HUMAN ETHICS APPROVAL

Human Ethics Approval was sought and granted from the University of Canterbury Human Ethics Committee for Phase One (the Pilot Study) and Phase Two (Experiments 1 to 4) of this thesis.